DOCUMENT RESUME

ED 191 324 FL 011 755

AUTHOR Culhane, P. T., Ed.: White, J., Ed.
TITLE Testing. Occasional Papers 21.
INSTITUTION Essex Univ., Colchester (England). Dept. of Language and Linguistics.
PUB DATE May 79
NOTE 119p.: Not available in paper copy because of small print size of original document. Some pages may be difficult to read in microfiche.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adults: College Students: Communicative Competence (Languages): Diagnostic Tests: *English (Second Language): *English for Special Purposes: Foreign Students: Higher Education: Immigrants: Language Proficiency: *Language Tests: Pronunciation: Sciences: *Test Construction: *Testing
IDENTIFIERS Air Traffic Controllers: England

ABSTRACT

Five papers on testing in English as a second language are included in this volume. "A Preliminary Diagnostic Test for Adult Immigrants" by Johan I. Arthur presents the first stage in a project to develop a test for limited English speaking adults in the Colchester, Essex area. "An Initial Testing Battery on a Course for Air Traffic Controllers" by C. Townley presents a description of that battery as well as a transcript of the tests themselves. In "Proposals for a Proficiency Test for Use on Pre-Sessional Courses," Lewis Kerr sets forth a test battery design suitable as an initial entry test for a pre-sessional course for foreign graduate students. Characteristics of the students are described, and the advantages and disadvantages of discrete item tests and integrative skills tests are weighed. In "Testing Communicative Competence in English as a Foreign Language with Students of Natural and Social Sciences," S. McDonough outlines reasons for the development of a new oral production test, and describes and analyzes a test procedure using graph material. Finally, K. Robinson, D. Rekart, and R. Daniloff offer "A Proposal: Use of Analysis by Synthesis to Investigate Naturalness of Second-Language Pronunciation," as a step in advancing the testing of foreign language pronunciation. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

University of Essex

Department of Language and Linguistics

OCCASIONAL PAPERS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P.T. Culhane, Univ. of Essex

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

October, May 1978

23

TESTING

Introduction

It is almost a decade since the publication of an Occasional Paper devoted to Language Testing. A comparison of the 1970 issue and the present one suggests that emphases within the field have changed. Then we were concerned with the validity of Language Aptitude Testing, specifically the 1959 Carroll Sapon MLAT, in the context of foreign language study in British universities. At that stage, MLAT was found wanting on two main grounds: its lack of any uniquely predictive power when compared with existing proficiency examinations, and secondly the problems posed by attempting to isolate a static aptitude factor from developing abilities in foreign-language learning. Perhaps the greatest contrast is that in the structure of MLAT certain factors are isolated which make assumptions about the way language will be taught. Methodology has greatly changed in emphasis, throughthe past decade and present teaching preoccupations are reflected in the types of test being proposed by the contributors to this issue.

Assuming that the papers collected here are representative of wider attitudes in the language teaching profession, the turning away from global, mass-survey approaches to language testing, is producing individually tailored tests, responsive to the needs of specific and often very small groups.

We can perhaps highlight the principal differences as follows:

1) There appears to be a closer relationship sought for between teaching and testing, evident particularly in the information-gathering interest of several of the tests.

2) There is renewed interest in 'the interview', as part of a movement away from wholly pencil and paper tests. (Of course, where literacy cannot be presupposed, such a development is obvious.)

3) There is a recognition that the expected final proficiency of L_2 learning varies according to the purpose for which that language is learnt, and so any test must be sensitive to this variation.

4) Relatedly, test-constructors are engaged in a more finely grained analysis of style and function in the language to be tested.

The emphasis that the following papers give these points comes in response to changing teaching methods over the past decade. As various authors note, discrete-item, structure based tests do not always fit theoretically with the prevailing interest in notional/functional syllabuses, nor with the groundswell of 'communicative' language teaching. The problems of establishing an appropriate test of so potentially nebulous a category as 'communicative language' cannot be minimised however. What kinds of texts should be used? What is an 'authentic' text? Are there ways of feasibly constraining

spoken language to make scoring possible without robbing (for example) conversational exchanges of all creativity? To what extent can a test of productive language in one domain (e.g. cloze or essay) be expected to correlate with other, perhaps untested, categories? How strong is the evidence for an underlying linguistic competence which will surface equally easily irrespective of the task set?

These questions are variously addressed here, and though a tendency towards integrative-skills tests is apparent, only one author proposes a single test as a theoretically viable replacement for the more familiar battery (see p.83). Circumstances again play a role: where the student's starting level in the language is low or virtually non-existent, a teacher/tester must cast a wider net initially (see p.1); again, where the course includes both specialist and general language skills, a range of tests seems more appropriate (see p.23). Nevertheless, one would like to see more research of this type on correlations, or lack of them, between newer techniques of testing and more traditional discrete-item tests. For while it is frequently attested that manipulative knowledge of decontextualised structures, words or phonemes does not presuppose fluent language use, the one set of correlations we have here does link success in 'productive communication' with that on a vocabulary and structure test.

One paper in this issue does in fact analyse the problem of specifying what is acceptably 'native' in the domain of pronunciation (see p.98). At present the proposals of these writers are theoretical, but given systematic research, they look forward to an accurate description of the crucial articulatory-perceptual characteristics of a 'natural' accent "within the decade".

The confidently exploratory tone of that proposal is repeated at least implicitly throughout this collection. It would be heartening to be able to anticipate a similarly co-ordinated development of the many existing proposals now current for testing different learner populations. We suggest that the ideas contained in this issue invite such collaboration.

P.T. Culhane

J. White

Editors

Occasional Paper 21

May 1979

Contents

		page
Johan I. Arthur	A Preliminary Diagnostic Test for Adult Immigrants	1
C. Townley	An Initial Testing Battery on a Course for Air Traffic Controllers	23
Lewis Kerr	Proposals for a proficiency test for use on pre-sessional courses	63
S. McDonough	Testing Communicative Competence in English as a Foreign Language with Students of Natural and Social Sciences	83
K. Robinson, D. Rekart, R. Danilloff	A proposal: Use of analysis by synthesis to investigate naturalness of second-language pronunciation	98

A Preliminary Diagnostic Test for Adult Immigrants

Johan I. Arthur

University of Essex Department of Language and Linguistics
Occasional Paper No. 21, pp.1-21

I Preliminary

This paper presents the first stage in a practical project, namely the development of a diagnostic test for adult immigrants with very limited English in the Colchester, Essex area. The next stage will be use of the test over several months and feedback on all aspects from the volunteer tutors administering it. A revised version will then be produced.

The orientation of this paper, as of the project, is practical, so that a description of context precedes and governs considerations of theory. Also, any conclusions must remain tentative in the light of the unfinished nature of the work.

II The Context

a) General (U.K.) Context

The population of modern Britain includes a sizeable number of immigrants, whose joint classification as a group belies the diversity, not only of their origins (notably the West Indies and the Indian sub-continent) but of their social and personal situations. The needs of this group, including their linguistic needs, cannot therefore be provided for by a monolithic programme: compare, for example, the needs of an unskilled factory-worker and those of a university-educated junior doctor, who have in common only that they have come, at different times and under different circumstances, from Pakistan. It is clearly necessary to be specific within any English teaching programme for immigrants, both in terms of aims and of a target group of learners.

BBC television, in its Parosi broadcasts, encourages Hindustani-speaking Asian women to come forward for free tuition by volunteers, a method of operation pioneered by the Adult Literacy Scheme. Unfortunately the needs of other immigrants have not been treated in like manner, so that the goal of relative learner homogeneity within the programme has proved unrealistic: it is not possible to turn away those who seek help simply because they do not fit bureaucratic categories. If students and their needs are so diverse, it follows that the learning syllabus, materials and also, where appropriate, assessment procedures, must be extremely adaptable. Given an individualised teaching situation, the concept of a battery of materials and also tests appear most useful.

The Parosi scheme, as indicated above, is aimed mainly at Asian women, who often lead very isolated lives while their husbands go out to work and their children, if they are old enough, to school. In this situation the volunteer tutor, who usually has no previous experience of teaching a language, not only has to teach everyday, useful English (in the light of the student's personal needs), but also has to be a link between the student and the host society. That is to say that the tutor, by befriending her student and building up her self-confidence until she is able to cope with various transactions alone, fulfils a social function.

Social aims cannot be said to be absent in the programmes of industrial language training certain firms provide for their immigrant work force. The main consideration here is, however, efficiency and the fact that "employees who cannot communicate competently in English cannot use the official network satisfactorily" (Fitzgerald, p.55). A more specialised, in fact company-specific programme is provided by ELT professionals.

Clearly the volunteer-staffed Parosi framework cannot operate along these lines. However, where firms, because of limited numbers of immigrant workers or for other reasons, do not give such

training, volunteers find themselves tutoring men whose linguistic needs are very different from those of the mainly home-oriented women described above.

b) Particular Context

In East Anglia, where the number of immigrants is comparatively small, volunteer tutors are called upon to help a wide range of students. The following brief profiles are illustrative:

1. Name: Mr. C.P.
   Age: 38
   Country of origin: Uganda
   Native language: Gujerati
   Length of residence in U.K.: 4 years
   Employment: Factory worker
   Family: wife (factory worker), three children
   Knowledge of English: poor

2. Name: Mrs. M.S.
   Age: 53
   Country of origin: Venezuela
   Native language: Spanish
   Length of residence in U.K.: 6 months
   Employment: none
   Family: daughter (nurse)
   Knowledge of English: none

The profiles, compiled from actual information about two students, not only illustrate their different circumstances and needs, but try to give an indication of their present proficiency in English. It is this undifferentiated and unsubstantiated indication, and means of rendering it more meaningful and informative, which will now concern us.

III Testing: Appropriateness and Inappropriateness

a) Purposes in Testing

Broadly speaking, it is useful to distinguish two purposes one might have for testing in the teaching situation we have described.

The first is that of diagnosing the students' weaknesses (which student and tutor can then work together to remedy) and, as Harris (p.3) points out, enables the plotting of "a performance profile which will show [...] relative strengths in the various areas tested." The obvious time for such a test is before or during the course of instruction and its obvious format a series of subtests (giving qualitative information about aspects of the student's performance).

A second possibility is the attainment test, administered during or at the end of a course in order to establish the efficacy or otherwise of instruction, both from the teaching and learning point of view. Such a test may be used for grading purposes i.e. assessing a student's progress in relation to other students and possibly placing him in a group on the basis of his grade. The role of tests or assessment of this kind is limited in the one-to-one teaching situation at present under discussion, although the student may, after a time, have the opportunity of moving on into a class or group, in which case placement could become relevant. A test of attainment can also be used as a means of encouraging the student and showing him the progress he has made; this can, however, best be done on an ad hoc basis, built into rather than supplementing or replacing normal lessons.

It is the test for diagnostic purposes with which we are here concerned. Tests of attainment will not be dealt with. Every effort should be made, in theory as in practice, not to confuse the two types of test (although, of course, some notion of level of attainment is implicit in diagnosis). Hill (p.35) points out the danger (in the context of teaching ESL in schools) of "useful diagnostic instruments ... used to establish attainment as a basis for dividing a group of children into streams, or to forecast future levels of attainment." The danger, of course, as with tests which claim to measure aptitude, is that of the self-fulfilling prophecy.

b) Methods of Testing

It has been established above that a student proficiency profile rather than a global assessment is required, and that a series of subtests, the results of which can be considered separately, is the most suitable format.

It would seem clear that the more rigidly formalised assessment procedures normally labelled "objective" are more able to give the kind of precise information useful in diagnosis, since the test situation can be made a tightly controlled one and specific structures etc. more easily elicited. Unfortunately the almost inevitable

isolation of items which facilitates checking also increases the artificiality of the situation. This artificiality, if it cannot be eliminated, should be minimalized for two reasons. Firstly, the needs of the adult immigrant obviously indicate a functional approach since what she can do using the English language takes priority over what she knows about it. Isolation of grammatical features etc. gives information primarily about linguistic knowledge. Secondly, it is necessary to strive, during what is probably the first meeting of student and tutor, for as relaxed and uninhibiting an atmosphere as possible; the apparatus of objective tests is likely to have the opposite effect.

The necessity of informal and, therefore, subjective testing is accepted, but objectivity and subjectivity form not a dichotomy but a continuum, so that one can and should attempt to standardise. Content areas can be largely laid down, for example, as the next section will show, and standard categories proposed for responses.

A test concentrating on functional use of language is, however, constrained by the fact that, as Heaton (p.82) points out, "success in communication often depends as much on the listener as on the speaker."*

The listener in this case is a volunteer tutor rather than an experienced language teacher practised in the skill of elicitation on the one hand and error analysis on the other. In terms of format and mode of administration this means that clear instructions and simplicity are essential; in terms of content it means that the diagnostic information gained must be comprehensible to and utilisable by the comparatively untrained tutor. (Short training courses are held, but are usually of no more than a few hours' duration).

---

* That is to say that the listener, as well as the speaker, plays an active role in the act of communication, bringing both linguistic and more general problem-solving techniques to bear upon the material to be decoded.

IV Test Content

It is often difficult in constructing tests to isolate and define precisely what one intends to measure. This is all the more so in the case of a functional approach such as that outlined above. Statistical validation is not a major concern here, however, and one should take the pragmatic attitude that content validity, as Pilliner (p.32) holds, "cannot be quantitatively measured but demands instead logical analysis and subjective judgement." To even claim that the "precise nature" of aspects of language can be defined and measured implies a certain view of language as a whole. Several views or approaches which have been found useful by language teachers and test constructors will now be discussed, and their appropriateness to the present testing situation evaluated.

a) Contrastive Analysis and Error Analysis

It is not within the scope of this paper to enter into the conflict between the claims of contrastive analysis and those of error analysis as explanations of learner language. Nor is the test designed to supply evidence for either hypothesis. Instead, both viewpoints are seen as valid and complementary. The fact that the students have different mother tongues is, of course, important and it is assumed that tutors will be given information on this during training and will be aware, for example, that most Asian languages do not have articles. The preliminary diagnostic test is concerned however with the individual, regardless of whether or not he fits such generalisations.

b) Language as Skills

A traditional approach to language teaching (and testing) distinguishes four language skills, namely listening, speaking, reading and writing. Clearly these skills are interdependent and it is not possible to isolate them satisfactorily. It is, nevertheless, useful to regard them as separable in the present context, especially since the exposure of immigrants to English is often far greater than their use of it, which is to say that their passive far outstrips their active knowledge. Also their opportunities for reading and writing may be severely restricted and, indeed, literacy in the mother tongue cannot always be presumed. In short, an individual's performance in the four skills may be very unequal. Given the conceptual familiarity to tutors of the traditional language skills, they may find it easier to detect such inequalities in these terms. Therefore, although the distinction is not and cannot be rigid, it informs the construction of the test. Primacy is given, in accordance with the needs of the vast majority of students, to speaking and listening i.e. oral communication. The test should, however, be flexible enough to allow some redressing of that balance in any particular case.

c) Elements in Language

Heaton (p.81) identifies the following components in the speech process: 1. pronunciation (including stress and intonation patterns).

2. grammar 3. vocabulary 4. fluency. (To this he adds a fifth component, that of comprehension, as a reminder of its central role in oral communication.) The usefulness of such categories in the test under discussion must depend on their contribution to intelligibility and communication, especially as students fall into the loosely-defined categories of elementary and intermediate rather than advanced. Social acceptability in speech cannot be entirely disregarded but must initially take second place.

Direct testing of vocabulary will be limited to functional areas in day to day activities (e.g. currency). Clearly, however, appropriate lexical use as a test of functional adequacy is a factor throughout.

Fluency, defined as ease and speed in speaking, is obviously a relevant factor in communicative efficacy. Its correlation with intelligibility is not, however, so obvious and both must be viewed as criteria.

Regarding pronunciation and grammar, one would ideally want a diagnostic test to isolate or pinpoint as far as possible faulty intonation, misuse of tenses, etc. The sophisticated test instrument this would require would also, unfortunately, require expert analysis of speech data (and imply subsequent expert use of the information gained). The volunteer tutor is not such an expert and would find such a test unwieldy and unhelpful. The task of analysis can, however, be made easier for the tutor by exposure to and discussion of learners' English in training session. Tape-recording the student, unless this is found to be too inhibiting, will also allow more accurate analysis.

Elements in writing which the test takes into account are limited to letter formation, which is important as many students are unfamiliar with Roman script and spelling. Tests of phonics etc. are available via the Area Organiser from the Adult Literacy Scheme operators, should further investigation here be required at any stage. Likewise material testing aspects of reading can be supplied. The reading required in the preliminary test is of an elementary nature and does not permit detailed analysis. This is in accordance with the main focus of attention on spoken language, as stated above.

V The Test

a) Guidance to Tutors

The Area Organiser will discuss the test with tutors before they use it and explain various points as follows.

The preliminary test should be regarded simply as part of the material at the tutor's disposal and as a useful format for the first lesson with his student. Its purpose is not to assign a grade or anything of that sort, but to bring out the student's strengths and weaknesses. On the basis of this the tutor can, with his area organiser's help as required, plan to cover certain areas, remedy certain mistakes etc. In addition, the test presented not as such but as a lesson, gives the student an opportunity to comment on what he finds difficult and would welcome help with.

Presentation is important, especially as priority in the first lesson must be given to establishing a friendly and relaxed working relationship. Nothing prevents the tutor, therefore, from using part of the test material one week and part the next. Nor need he insist upon the student's completing a section found bewildering or irrelevant. In that case a comment can be made on the Findings Sheet and the tutor can possibly have a word with his Area Organiser.

The Findings Sheet is designed to serve as a reminder and help in subsequent lessons, and can also be handed on to another tutor if, for any reason, there is a change. Like the rest of the test, it is far from perfect, being something of a "pilot project", and so comments, critical and otherwise, will be more than welcome. The Area Organisar will arrange to talk to each tutor at some point and find out what he thinks about the areas the test covers, level of difficulty, presentation, usefulness of the information it provides etc. A revised and improved version can then be prepared.

b) Tutors' Test Sheet

NB's: Remember to give your student lots of encouragement: keep corrections to a minimum, especially at first. Timing is of no consequence, although you should break off and go on to another item if it is clear the student is making no headway at all in any section.

The order of sections can be re-arranged if it seems more appropriate or logical to do it that way.

Section A

Conversation: giving information.

The following questions should be put to the student as naturally as possible. Although they will be the only ones considered for the Findings Sheet, you can include others that crop up or ease the conversation along. Questions can also be rephrased (simply) to make them more natural in the flow of conversation.

1. How are you?
2. What's your name?
3. Can you spell it for me?
4. Where do you live?
5. Where do you work?
   (or What do you do at home?)

Section B

Conversation: asking for information.

The student is to be encouraged to ask the tutor questions similar to those in Section A. Begin by saying: How can you ask about me? or What would you like to know about me? Prompt as necessary so that the student attempts to form questions about

1. Your name
2. Where you live
3. Where you work
4. Your family

Section C

Basic writing: letter formation.

The student is to be asked to write down his name and address on

the student profile form supplied by your Area Organiser. If he has no difficulty he can go on to fill out the rest of the form. Otherwise he should be given a piece of paper onto which to copy your name and address for his future reference.

Section D

a) Numbers

A dozen or so pencils, counters or any other suitable items will allow you to ask: How many .... are there here? and how many now? et cetera.

b) Telling the time

If the student has shown he can cope with numbers, a toy clock or an old real one will allow you to ask: What time is it now? etc. For the student who is not at this entirely elementary stage, questions can be asked instead about her daily routine, e.g. What time do you get up in the morning? Check on the phrases:

1. ..... o'clock
2. ..... past .....
3. ..... to .....
4. ..... quarter past .....
5. ..... quarter to .....
6. half past .....

c) Money

Again assuming he can cope with numbers, £1 in change, plus a £1 note is possible, can be used to ask the student to name:

½p (pence or pee), 1p, 2p, 5p, 10p, 50p, £1

Now put some money in a pile and ask the student how much there is. Repeat two or three times.
N.B. With a student who has passed the elementary stage it will again be better to ask directly about prices, fares he pays etc.

d) Dates

A quick check can be made of whether the student can

1. Give today's date (day and month)
2. Name the other days of the week
3. Name the other months of the year.

Section E

Following simple instructions.

Any items handy on the table can be used and the student told vocabulary for items he cannot name. Avoid give-away gestures. The student should be asked to

1. Pick up the book
2. Open it
3. Put it down next to the pen (pencil etc.)
4. Now pick up the ruler
5. Put it on top of the book
6. Give me the book
7. (Handing the book back open). Shut it, please.

Section F

Recognising situational conversation.

The short conversational exchanges on the accompanying tape should be played, one at a time, to the student. After each exchange, the student should be asked to point out the appropriate scene in the series of pictures provided (the student will need a chance to familiarise himself with these before beginning.)

Ask him: Where do you think those people are?

1. Oh, and I've two letters for Europe.
   How much will they be?

   Ten and a half p. each. Is that all?

2. You can't park here, you know.

   Why not? I can't see any yellow lines!

3. Good morning, doctor, I'm not too well, I'm afraid.

   What seems to be the matter, then?

4. When's the next train for Colchester, please?

   Not for another hour. You've just missed one.

5. Sugar's up this week. Do you want any?

   No thanks, not at that price. But give me a packet of those biscuits, please.

Section G

Pronunciation and intonation: repetition and reproduction of a conversation.

The following short taped dialogue is to be played twice to the

student and any necessary explanations given. Tutor and student should then take turns at reading the two roles and finally the student should be encouraged to reproduce freely. Note that this is not a test of exact memory or speed of learning, but rather an opportunity for the tutor to concentrate on features of the student's pronunciation and intonation.

In the Grocer's Shop.

Customer: Good morning. Give me a pound of butter and a dozen eggs, please.

Assistant: Certainly. Is there anything else.

Customer: No, thank you.

Assistant: That'll be one pound ten, please.

Customer: I'm afraid I only have a five pound note. Can you change it?

Assistant: That's all right - I think I can manage. That's three pounds ninety change.

Customer: Thanks very much. Bye!

Section H

a) Basic reading

The following sentences are reproduced on cards for presentation to the student. He should be asked to arrange them in the correct sequence.

1. At eight o'clock every morning Mr. Smith gets up.

2. He has breakfast and a cup of tea.

3. He leaves the house and begins work at nine o'clock.

b) Basic writing (spelling)

The teaching cards should be set aside and the three sentences dictated clearly and as slowly as required to the student.

Section J

Talking about a picture (comprehension of tenses).

The following questions refer to the picture "The Accident" and are designed to include various English tense forms. The student's answers will indicate whether he understands time references. The

questions can be included in a more general chat about the picture. Vocabulary should be given as required.

1. What is the policeman doing?
2. Where is the car?
3. Why has the ambulance come?
4. What are the ambulance men going to do?

Also, if the student has coped well:

5. Have you ever had an accident?
6. What would you do if you saw an accident like this?

c) Findings Sheet

Tick off findings, adding comments as necessary

(Note: acceptable = understandable and appropriate)

Section A		Section B	
responses:	acceptable unacceptable	questions:	acceptable unacceptable
1		1	
2		2	
3		3	
4		4	
5		any others	

Section C

unable
block capitals
slowly, with difficulty
with reasonable ease
no difficulty

Particular letters or features causing difficulty:

Section D

a) Numbers known: not at all
a few
most
all

b) Able to use 1
2
3
4
5
6

Numbers causing difficulty:

Section D continued

c) Currency known: not at all
some
most
all

particular difficulties:

d) days known: not at all
some
most
all

particular difficulties:

months known: not at all
some
most
all

particular difficulties:

Section E

Instructions carried out correctly:

1
2
3
4
5
6
7

Section F

Situations recognised correctly:

1
2
3
4
5

Section G

Difficulties noticed in pronouncing particular vowels (which?)

consonants (which?)

Difficulties (or inappropriateness) noticed in intonation patterns of questions?

statements?

greetings etc.?

Section H

a) Able to order cards correctly:

yes

no

b) unable

gross, random spelling errors

many errors but showing some phonic consistency

few errors

no difficulty

Section J

Appropriate response to: (showing comprehension of tense)

1

2

3

4

5

6

VI Comments on Test Sections

The test sections will now be considered in logical groupings from the point of view of purpose and form.

Sections A and B: Social survival - interaction.

These sections separate the absolute beginner from the elementary-stage student. The information gained is whether or not the student has basic linguistic means of social survival. Additionally, Question A3 is a quick and natural way of ascertaining knowledge of the English alphabet.

Section D: Social survival - functional areas.

This section, parts of which can be omitted or delayed, establishes the student's degree of competence in areas to be covered by a basic syllabus tailored to his or her needs and thus indicates the extent to which they should be concentrated on.

Sections E,F and J: (Listening) comprehension.

The first two of these sections (E and F) test comprehension at a survival level, and do so without the need for verbal responses on the student's part. Section J is more difficult and attempts to establish the student's comprehension of specific tense forms. Here a verbal response is required, but comprehension is not judged entirely by the accuracy with which the student reproduces any tense form. The judgement made is, therefore, largely (and unavoidably) subjective.

Section G: Spoken skills.

This section provides an opportunity for establishing the degree to which the student's pronunciation and intonation approximate to those of native speakers. Ideally the student can be recorded: the repetition involved will, however, facilitate on-the-spot analysis.

Section C and H: Reading and writing.

Sections C and Da) test reading and writing skills at a basic survival level. In addition to the analysis on the Findings Sheet, the student's script from Sections C and Db) can be kept for future reference as the subsequent teaching syllabus takes shape.

It is hoped that purpose and form of the Test will fulfil the dual aim of the introductory lesson as formulated in a handout to tutors:

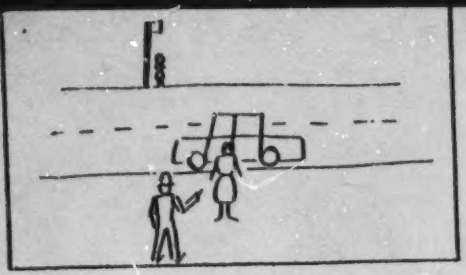
(I) To get to know the student and to establish friendly relations.

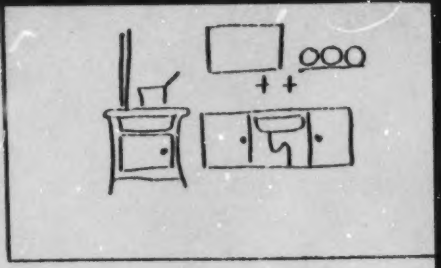
(II) To discover areas of difficulty and knowledge of English for planning future work.

VII Conclusion:

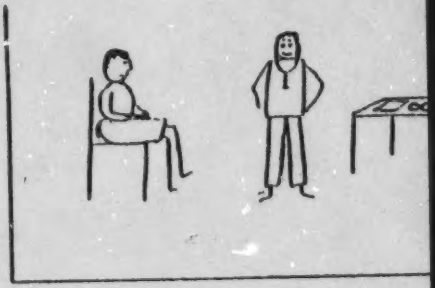
Areas for Further Research and Development

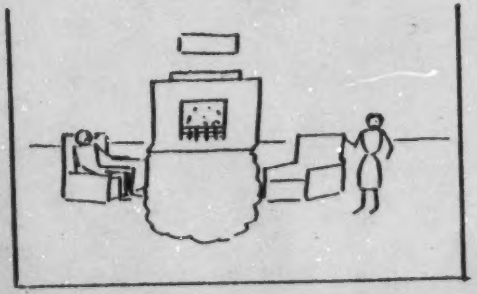
The test outlined still has to prove its usefulness and effectiveness. This will be established via feedback from tutors and students. It will probably also be important for the revised test, as well as incorporating this feedback, to be made more versatile and thus suitable for use with students who have progressed beyond the survival level on which several test sections are based. To this end supplementary (optional) sections need to be developed. These may eventually be part of a battery of subtests, for use at any appropriate time, corresponding to the battery of materials necessary in individualised tuition. The concept of an initial diagnostic test will probably be retained, however, not least for the support, both practical and psychological, it gives to the new volunteer tutor.





POST OFFICE
LETTERS





THE ACCIDENT

APPENDIX A

References

Fitzgerald, M. (1977) "Industrial Language Training" in Holden, S. (ed.) English for Specific Purposes, MET Publications.

Harris, D.P. (1969) Testing English as a Second Language, McGraw-Hill

Heaton, J.B. (1975) Writing English Language Tests, Longman.

Hill, D. (1976) Teaching in Multiracial Schools, Methuen

Ingram, E. (1974) "Language Testing" in Allen, J. and Corder, S.P. (eds.), Edinburgh Course in Applied Linguistics, Vol.3, OUP.

Jupp, T.C. & Hodlin, S. Industrial English, Heinemann

Parosi Tutors' Handbook (1977) BBC

Pillinor, A.E.G. (1968) "Subjective and Objective Testing" in Davis, A., Language Testing Symposium, OUP.

Sealey, G. and Skirrow, G. (1972) Scope (Teachers' Book, Senior Course), School Council/Books for Schools Ltd.

An Initial Testing Battery on a Course for Air Traffic Controllers

C. Townley

University of Essex Dept. of Language and Linguistics
Occasional Paper No.21, pp. 23-61

SECTION I: INTRODUCTION

1. The School. The Royal Air Force has a limited commitment to provide English courses for overseas students, officers and airmen from foreign and commonwealth air forces, who come to Britain on specialised training courses. The professional training includes officer staff college courses and courses in airmanship, air traffic control and electrical engineering which are held at specialised units throughout the country. English teaching takes place at a small unit, English for Overseas Students (ELOS), within the School of Education, at Royal Air Force Newton, near Nottingham. This unit has a staff of four education officers, with a mixture of language training and technical experience, plus clerical support. It also shares the excellent training facilities available in the School of Education, which include a language laboratory and a wide range of audio-visual equipment as well as a comprehensive training resources support.

2. The Air Traffic Control Course. All overseas students arrive with a specified minimum level of English having undergone a selection procedure administered by the Foreign and Commonwealth Office. The ELOS unit has no part in this selection system and in practice standards of English vary considerably amongst students. The particular course which this paper focuses on is the English for Air Traffic Control course. This is an 8 week course the aims of which embrace, very broadly, the following main elements:

a. General Remedial. The course aims to provide general EFL teaching to improve English standards for general and social purposes.

b. Subject Specific. The main ESP aim of the course is to improve the students' knowledge of general aviation terminology and to introduce the language of air traffic control. A secondary aim under the heading of ESP is to help prepare students linguistically for the academic demands of the professional training course that follows.

3. An English for Air Traffic Control (EATC) course typically has 6 to 8 students. Subject to normal service diversions, such as sport and administration, which effectively reduce the training week to 4½ days, the course comprises 8 weeks of full-time, intensive language training. The training programme and methods cannot be discussed within the confines of this paper, but it must be said that the training specification is based on a thorough needs analysis carried out in conjunction with the ATC school, and there is regular liaison between the two schools.*

---

* Practically all training in the Royal Air Force is based on a closed loop systems-approach-to-training (SAT) to ensure that training does in fact meet operational needs.

Courses are provided for students from East Africa, West Africa and several Arabic speaking countries. It has not been feasible, therefore, to use contrastive analysis in the construction of the test battery that forms the subject matter of this paper. If the battery proved successful in practice, however, it might later prove possible to modify sections of the tests to accommodate linguistic differences among the students. Whether the battery as it is presently envisaged gives any advantage to any particular culture group can also only be determined in practice.

4. Aims of the Project. An underlying aim of this project is to consider to what extent English language testing can promote the effectiveness and efficiency of language teaching in the school. This clearly views testing not as a distinct and separate part of teaching, but as an integral component in the whole training process. The more specific aim is to outline the overall structure of an initial language test battery to assess students' proficiency in productive and receptive linguistic skills. The tests of oral, writing and reading skills will be given only in outline to allow an overall perspective. The aural tests, which comprise the major part of this paper, are given in full.

5. Format. The format of the rest of this paper is as follows. Section II considers the scope and aims of the test battery. It briefly discusses how testing may be viewed as an integral part of training and goes on to describe the criteria used in the selection of test material. Section III gives an overview of the test battery by giving an outline of the format and content of the first 3 parts of the battery. Section IV presents a more detailed commentary on the listening comprehension tests, which, for convenience, are attached as a separate supplement at Appendix 'A'. The final section is a brief summary of some of the main points contained in this paper.

SECTION II: SCOPE AND AIMS OF THE TEST BATTERY

6. The test battery consists of 4 groups of tests which are designed to serve both as proficiency tests and as diagnostic assessments for training purposes. As a measure of linguistic proficiency the battery is intended to serve two purposes:

a. External Validation. Tests should monitor intake standards of students over time, thereby providing a check on the effectiveness of the selection system conducted by an outside agency.

b. Internal Validation. Testing should form the first step in the internal validation of training. This clearly requires subsequent testing for comparative purposes, and thus could form one evaluation of training to be used in parallel with the external validation of training to be carried out in conjunction with the ATC course and later feed-back from operational units.

7. From a purely training point of view the more important aim of the test battery is for diagnostic purposes. It is envisaged that, while parts of the ESP elements of the course would continue to be taught by a variety of group teaching methods, remedial EFL teaching could be carried out on a highly individualised basis on the results of initial diagnostic testing (and continuing assessment). To be effective such individualised tuition would depend upon its success in meeting 3 fundamental criteria:

a. Assessment Criterion. An accurate assessment must be made of the particular linguistic needs of the students concerned, and each student's proficiency under each identified linguistic heading.

b. Recording Criterion. Detailed records must be drawn up of each student's various proficiency levels and regularly amended to assess his progress.

c. Remedy Criterion. Appropriate and adequate materials and/or instruction must be available to meet the needs identified for each individual student.

8. From this perspective, then, testing is seen, not as an end in itself, but as an essential and integral part of the whole training process.* Of the 3 criteria, the first, it is hoped, will be met by this test battery, including the assessment of linguistic needs which is considered later. The second, the recording criterion, will require the construction, completion and maintenance of a detailed student profile, the first input of which would be the results of the initial test battery. The feasibility of maintaining detailed and comprehensive student records must depend, inter alia, on the student to staff ratio, and in some circumstances it may well be impracticable (Lee 1957). With so few students as there are on the FATC course, however, and in view of the specific aims of the course, it is not only feasible but also necessary. Further discussions of the student profile would take us outside the scope of this paper.

9. The third criterion requires appropriate remedial measures, in whatever form, to be available to meet students' needs. Again this merits only brief mention within the confines of this essay, but at least 3 sources are envisaged. The staff are available to give extra work and help with particular problems. Peer instruction is of great importance on a course of this kind and group cooperation must be

---

* It must be stated that I have considerable sympathy with the minimal teaching strategy advocated by Allwright (1976) and see potentially useful applications of this in elements of the ESP course where there is a need to foster group independence in preparation for subsequent ATC training. On other parts of the course I regard teacher guidance as both necessary and desirable and it is here that effective monitoring of progress is an important factor. As has been argued elsewhere (e.g. Long 1976) minimal teacher intervention need not be viewed as an all-or-nothing approach but can form a useful part of a more conventional method.

encouraged. Finally, self-access remedial materials in modular form may be the best solution for some linguistic difficulties.

10. We turn now to consider the reasoning behind the scope and content of the initial test battery. This we do by looking first at the gross linguistic skills requirements of the ATC students. For operational purposes (and it would unnecessarily complicate the issue to try to take into account EAP and general EFL needs) the students' linguistic needs can be allocated the following relative importance weightings as amongst the 4 skills:

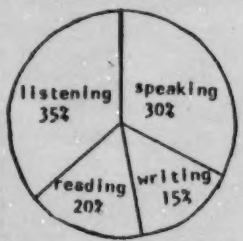


Fig.1: Assessed relative importance of language skills in ATC.

This for the present is a subjective assessment and the figures may require future amendment. As a start, however, they give some basis on which to work. The reasoning behind the breakdown is as follows. Writing is seen as the least of the controller's linguistic needs. Most of the writing he is required to do is in the course of administrative rather than operational duties. Reading merits a higher rating in that the controller must be able to read flight plans and current-state visual displays. These, however, are in standardised format and their linguistic range is very limited. The two key skills, then, are oral and aural; his main responsibilities entail receiving air/ground communications and transmitting ground/air information and instructions. In all routine operations the linguistic range is fairly limited, both orally and aurally, and both skills might merit similar importance ratings. In emergencies and other non-routine operations, however, extension of the linguistic range is more probable in air/ground communications than in ground/air; that is to say that the range of emergency procedures to be followed in the ATC tower is more narrowly prescribed than the possible events which could occur in the air. For this reason the listening skill is given a somewhat higher rating than the speaking skill.

11. The second major dimension along which the scope and content of the test battery may be considered is in terms of the linguistic elements it seeks to test. Although it is presumed that the students require a specified threshold level across the entire range of language, it is clearly not possible to test so widely, not at least in specific terms. It is intended, therefore, in different sections of the test battery to focus on selected linguistic elements so that the battery in aggregate covers all those features which have been identified as of particular importance to the air traffic controller. The way in which the test battery attempts to give adequate coverage both of the language skills and the linguistic elements is shown in

Fig.2 on page 28. Once again it must be admitted that the assessment on the relative importance of linguistic elements has been largely subjective, but again it serves as a starting point for the rational and systematic identifications of training needs even though it may require future amendment.

SECTION III: OVERVIEW OF THE TEST BATTERY

INTRODUCTION

12. The initial test battery as envisaged is considerably longer than most similar proficiency tests; in all it requires almost 4 hours to administer, though this is spread over the first week of training. It is considered that the length of the tests is justified by the nature of the intensive course and the integral part that testing is seen as playing in that course. Furthermore, some of the tests, such as the interview part of the oral test, are integrated into and take advantage of normal arrival procedures which are part of the routine initial administration on any Service unit. Thus the interview merely formalises a form of assessment that would occur anyway.

13. The tests may be seen as falling into 3 broad categories:

a. Tests that focus specifically on certain linguistic features. For example, Test 1D, Aviation Description, focuses specifically on the student's knowledge of basic aviation lexis, while Test 4E in the listening comprehension is concerned only with the student's knowledge of English numerals.

b. Integrated tests which seek to assess a student's proficiency in one of the gross skills. Examples of this are the interview and discussion in the oral part of the battery.

c. Tests that tend to focus on specific language elements while remaining basically integrated skill tests in nature. An example of this is Test 2A, a written narrative that will tend to focus on the use of past rather than present tense verb forms.

14. Tests may also be distinguished in the manner of their assessment. In the oral and writing tests it has been accepted that subjective assessment cannot be avoided in all areas. It is intended, however, that any possible disadvantages of this will be offset both by the employment of systematic recording methods, and, where possible, by the use of more than one assessor. In the reading and listening tests assessment is almost entirely objective. In the remainder of this section we present an outline of the first 3 parts of the battery and discuss some of the criteria that underlie the nature and content of the tests. An overall view of the test battery, showing skills areas and linguistic elements, is shown in Fig.2 on page 28.

tests			PHONOLOGY				LEXIS						STRUCTURES							
			stress	intonation	phon.discrim.	enunciation	gen. vocab	aero. vocab	idiom	numerals	time	spatial	verb forms	modals	questions	negatives	pronouns	prepositions	adject. forms	adverbs
		importance rating**	3	3	4	4	4	5	3	5	5	5	4	4	4	4	3	5	4	3
SPEAKING	1A	reading	*	*		*														
	1B	interview	*	*		*	*		+				*			*	*	*	+	+
	1C	discussion	*	*		*	*		+				*	+		*	+		+	+
	1D	aero.descrip.	*	*		*	*	*					*					+	+	
	1E	gen.descrip.	*	*		*	*		+				*						+	
	1F	presentation	*	*		*	*		+				*							
WRITING	2A	narrative					*		+				*			*	*	*		+
	2B	description					*		+				*						+	
	2C	discussion					*		+				*	+						
READING	3A	gen.vocab.					*													
	3B	aero.vocab.						*												
	3C	tenses					+						*			+				
	3D	modals						+						*				+		
	3E	questions													*					
	3F	adjectives					+												*	
	3G	prepositions					+						*					*		
	3H	reading comp 1					*	*	*		*		*			*			*	
	3I	reading comp 2					*	*					*	*				*		
LISTENING	4A	phon.discrim.			*															
	4B	synonyms					+							*					*	
	4C	responses												*		*				
	4D	route map					*						*					*		
	4E	numerals								*										
	4F	time									*									

* : test focusses on particular linguistic features

\+ : test will probably highlight particular linguistic features

**:Note: Importance ratings are on a scale from 5 (essential) to 1 (unnecessary)

Fig.2. A BASIC METHOD OF ENSURING TEST COVERAGE OF LINGUISTIC NEEDS IDENTIFIED BY NEEDS ANALYSIS.

The data in Fig.2 is to some extent subjectively selected. It is, however, in no way intended to be definitive or exhaustive, particularly in the area of grammatical structures. It merely illustrates the need to:

a. identify the important linguistic elements required by a particular group of ESP students, and assign a weighting to those elements.

b. identify the linguistic skills required by those students and assign an importance rating to them.

c. view these two factors together in order to validate the content of the training syllabus or test battery.

TEST 1: ORAL PROFICIENCY

15. The oral tests consist of 5 sub-tests lasting almost an hour, and spread over 3 days. They are:

a.	Test 1A: Reading aloud	2 minutes	Day 1
b.	" 1B: Personal interview.....	20 "	Day 1
c.	" 1C: Topical discussion.....	10 "	Day 2
d.	" 1D: Aviation description.....	10	Day 2
e.	" 1E: General description.....	10 "	Day 2
f.	" 1F: Student presentation	5 "	Day 3

16. Reading Aloud (1A). This 2 minute test is intended to serve as a brief ice-breaker before the interview starts, and as a simple means of identifying from the very beginning any major problems of stress, intonation or pronunciation. The student reads aloud one or more passages of intermediate English, having been given the opportunity to read through them for 10 minutes before he came in.

17. Personal Interview (1B). As a matter of normal administrative routine, a student arriving on a training course receives an in-depth interview with his flight commander. An overseas ATC student would thus be encouraged to talk about his family background, education, career to date, first impressions of Britain, personal problems and so on.. With a systematic method of recording the language content of this interview it can be productively used for linguistic analysis and initial assessment. Tape recording is perhaps not a good idea at this stage, as it is essential that the interview takes place in a relaxed and informal atmosphere, in which the student can be encouraged to express himself freely. Two members of staff are needed at this interview, one to conduct the interview, the other to make notes on a prepared assessment proforma. Assessment at this

stage must, of necessity, be gross, but it should be possible to evaluate on a 1 to 5 scale overall oral ability, plus some broad linguistic features under the headings of phonology, lexis and grammar.

18. Topical Discussion (1C). The second group of oral tests takes 25 minutes and is scheduled for day 2. These three tests may be categorised as "general - tending to focus" as described above. The first of these is a discussion on a subject of topical interest lasting 10 minutes. One perhaps useful way to select the subject matter for this would be to give to the student half an hour before the session the front page of the Times or Telegraph, from which he can select one or more items of news that he would be prepared to discuss. The setting for this test and the two that follow should be similar to that used in the interview - a relaxed, informal atmosphere with two members of staff, one recording on a prepared proforma.

19. Aviation Description. (1D). This 10 minute test is intended to give an insight into the student's knowledge of basic aviation terminology. The test is based on a booklet composed of 10 pictures depicting aircraft, common equipment and aviation scenes. The student is asked to describe what he sees in each picture so as to give, primarily, a subjective assessment on a 1 to 5 scale of his aviation lexical knowledge (plus whatever phonological and grammatical spin-off occurs).

20. General Description (1E). This is precisely the same kind of test as 1D but in this case the pictures in the booklet are of a more general nature. Suitable topics might be transport, industrial scenes, high street scenes and so on. It is important only that the student is given enough opportunity to use his lexical knowledge. With tests 1D and 1F it would again be desirable, though not essential, to have two members of staff so that one could concentrate on tape recording.

21. Student Presentation (1F). The first 6 oral tests have all been conducted on an individual basis. The seventh, the student presentation, requires the student to speak formally in front of his fellow students for the first time. For this he requires at least 2 days warning to select his topic and prepare his material. The presentation, 5 minutes long, may be on any topic selected by the student (from a list of suggestions if necessary). The only clear instruction necessary is that the presentation should not be written out in full and read aloud. Students should be encouraged to keep their notes fairly brief; it matters little whether they are in English or their mother tongue.

22. It may be argued that some of the oral tests are impure and this is quite true. Test 1C and 1F both require at least some reading and writing ability. From a testing point of view, however, it seems of little importance when assessment is subjective and necessarily gross in nature. Furthermore, these subjective tests also form an important part of the training content of the first week of the course and play a considerable part in getting students to talk freely. It may also be said that these are intermediate students and there are many real-life situations in which it would be artificial to try to

divorce completely the language skills. Thus few native speakers would attempt to give a formal presentation without doing some preparatory reading and writing some notes.

TEST 2: CREATIVE WRITING

23. Some basic aspects of the writing skill might have been tested by objective means, such as spelling, punctuation and selection of correct grammatical forms. However, since the only concern is with communicative ability so that technical factors such as spelling and punctuation are only of importance in the extent to which they impede communication, and since many of the grammatical forms are adequately tested in Tests 3 and 4, it was decided that open ended writing assessment was preferable. Thus, though the 3 tests of writing are subjectively assessed, they offer the possibility of much deeper insights into the students' linguistic competence than could be achieved by objective testing. With so few students it would be possible, and perhaps useful, to have each written test marked separately by two assessors.

24. Test 2 consists of three tests taking one hour to complete:

a. 2A: Narrative ..... 20 minutes.

b. 2B: Description .... 20 minutes.

c. 2C: Discussion ..... 20 minutes.

25. Narrative (2A). The student writes a short narrative from a choice of topics all of which are selected to allow him to draw as far as possible on his own experience. This is a general writing test but should tend to focus on to past verb forms and perhaps prepositions of time and personal pronouns.

26. Description (2B). In this test the student writes a short description, for example of his home town or some other place he knows well. This too is a general test, but may tend to focus on present verb forms, prepositions of place, and adjectives.

27. Discussion (2C). The student writes a brief discursive passage on a topic of current debate - perhaps on some basic aspect of science, technology or defence. It is more difficult to anticipate any tendency to focus in this case, but it could allow use of modals and impersonal verb forms. It is probably better to regard this as a purely general test of written English.

TEST 3: READING COMPREHENSION

28. Although the controller's key skills have been defined as speaking and listening, the tests of reading comprehension allow most objective insights into the student's knowledge of linguistic structures. For this reason the reading tests are the longest in the battery, lasting a total of 69 minutes. They may be categorised as follows:

a. Tests 3A and B. Tests of general and aviation vocabulary lasting 15 minutes.

b. Tests 3C-3G. Tests of selected linguistic structures lasting 30 minutes.

c. Tests 3H and J. General integrated tests of reading comprehension, requiring 20 minutes.

Very brief outline of each test, with an example of the content, are given below.

29. General Vocabulary (3A). This is a 10 minute test with 40 multiple choice items.

Eg. furious.....	a. strange b. harmful c. in flames d. angry	consume....	a. eat b. believe c. start again d. smoke

30. Basic Aviation Vocabulary (3B). This is a 5-minute modified cloze test - an account of a jet trainer taxying out, taking off, making a short cross-country flight, rejoining the circuit and landing. Aviation terms are omitted at selected points and the student has to select an appropriate word from a choice of 4.

Eg. "..... a. altitude
b. pressure
.....c. range is 21,000 feet
d. direction

and I am descending ....."

31. Verbal Tenses and Auxiliaries (3C). This is a 10 minute test with 30 items. The first 15 of these are multiple choice:

Eg. "This runway ..... for more than 5 years now."

a. is not repaired
b. was not being repaired
c. has not repaired
d. has not been repaired

The second 15 are not multiple choice but require the student to give the correct form of a verb in context:

Eg. "I (KNOW) him when he was training."

Eg. "If the rain had stopped earlier more people (COME) to see the parade."

32. Modals (3D). A 5 minute multiple choice test with 15 items:

Eg. "Before landing the pilot ..... to check that his wheels are down and locked."

a. has b. must c. wants d. should

33. Interrogatives (3E). The test of interrogatives takes the form of 12 short questions and 16 short answers; to each question there is one, and only one, correct answer. Four of the answers are redundant. The student has to match the letters of the questions with the figures of the appropriate answers. The test is given in full:

A. When did you arrive?	1. No he doesn't.
B. How long are you staying?	2. From London.
C. Where are you going?	3. It's late.
D. Where is your home?	4. My friend told me.
E. How old is your brother?	5. Only yesterday.
F. He doesn't sing well does he?	6. About two weeks.
G. What did you buy?	7. To London.
H. How much was the book?	8. Almost 18.
I. Why are you leaving?	9. Three thousand metres.
J. Isn't it warm?	10. Yes, she is.
K. How do you know?	11. One pound fifty.
L. How long is the runway?	12. Yes, it is.
	13. Near London.
	14. Yes it isn't.
	15. A book.
	16. Every Thursday.

34. Adjectives (3F). This is another 5 minute test with 15 multiple choice items on comparative and superlative forms of adjectives:

Eg. "The River Volga is the ..... river in Europe."

a. longer
b. most long
c. most longest
d. longest

35. Prepositions (3G). This is a 5 minute multiple choice test with 15 items focussing on prepositions of time and place:

Eg. "He arrived in Britain a. at b. on c. in d. until October last year."

Eg. "There are many beautiful villages a. between b. among c. from d. through here and London"

36. Reading Comprehension (3H). The reading comprehension is the first purely general test of the integrated reading skill. It is based on a passage with an aviation flavour (but containing no difficult technical vocabulary) with 6 multiple choice questions.

37. Reading Comprehension (3J). In the second reading comprehension the student is given a passage which describes the flight of a light aircraft between two points and an outline map. On the information given in the passage he has to trace the route of the aircraft on the map. The information will be given in the form of spatial prepositions, directions and salient features, such as rivers, mountains, railway lines and large buildings. (The reading comprehension tests take 10 minutes each.)

SECTION IV: LISTENING COMPREHENSION TESTS

INTRODUCTION

38. The auditory tests of the battery are intended to provide the initial input to the student profile on the integrated listening comprehension skill plus two linguistic elements that have been identified as of particular importance to air traffic controllers and which are effectively tested by auditory means. The tests in this section therefore consist of:

a. One test of phoneme discrimination.

b. Three tests of listening comprehension

c. Two tests focussing on particular linguistic elements.

PHONEMIC DISCRIMINATION (4A)

39. In the phonological field there were three main areas open to testing: sensitivity to stress, to intonation and to phoneme differences. Each of these was considered in the light of the students' needs. Tests of stress awareness were rejected on the very subjective grounds that, in isolation, they seem to serve no particularly important purpose. Identification of tonic stress is certainly a subconscious process that must occur in aural comprehension and a desirable feature for oral intelligibility. On the other hand, it is conceivable that one could pinpoint the tonic in a completely unknown language without this in any way assisting comprehension. For the needs of EATC, therefore, it was decided that the testing of stress in isolation would be a nugatory exercise.

40. A test of intonational awareness was regarded as much more meaningful - in the proper circumstances. EATC, however, does not constitute the proper circumstances. Intonation is the subtle manipulator of mood and meaning; it is frequently used effectively in irony, for example. In operational ATC communications, however, this kind of subtlety is uncommon - the usual quality of communication militates against it. For this reason no test of intonation was included in this section.

41. If all the ATC students came from the same linguistic background it would have been easier to decide whether or not to include a test of phoneme discrimination. Any test then might have been scientifically based on contrastive analysis. The fact that the students have different backgrounds made it a more difficult decision. The test that has been included is something of a compromise. It is a fairly short test of 40 items (of the 100 items in Ingram et al) whose intention is not one of delicate analysis, but only to serve as a gross measure in order to identify ab initio any major problem areas.*

42. Several factors influenced the choice of form and content of the test. Pictures were considered in various forms as a means of keeping the test "pure" but with students from at least 5 different cultural backgrounds the problems of illustrative conventions would have been too great. As observed in Lado (1961) "pictures are an excellent way of testing auditory comprehension but they are definitely limited and they can be thoroughly ambiguous." Another relatively pure test method is that of Davies (1964) in which 3 words are spoken and the student has to decide which, if any, of the words were the same. This requires no reading or spelling ability, but it does suffer from a rather ponderous system of recording the answers which makes it somewhat complex. Had the students concerned been beginners then this method would have been preferred to pictures.

43. The method actually selected was the underlining of one of three printed words as used, interal, by Ingram. This method has some disadvantages, but was considered most suitable to the needs of EATC students. One obvious objection to it is that it is impure in that it demands also a reading ability. The students, however, are (hopefully) at an intermediate level of English, and the content of the test was furthermore kept fairly easy. Another shortcoming of the selected method (and many other methods) was the additional difficulties caused by decontextualisation. To have contextualised each item, however, would have made the test much longer in time than it deserves to be. It can be argued also that air/ground communication, though not decontextualised, contains far less redundancy than is normal and messages tend to be very succinct. Partly to offset the difficulty of decontextualisation, however, each word in the test is read out twice, with a short pause in between.

44. The test as set, then, comprises 40 items requiring 8 seconds per item, a total of 7 minutes including instructions and examples. The first 20 items seek to test some of the more common vowel sounds of English; all words are monosyllabic and each group of three differs

---

* Lado (1961) quite rightly says that ".... we should ideally have separate tests for the various native language backgrounds of the subjects." In the case of the EATC course this in fact may turn out to be feasible, particularly in the area of phoneme discrimination.

phonemically only in its initial or internal vowel. The second half of the test covers a selection of consonant distinctions including some consonant clusters. Differences are initial, internal and final. Heaton (1975) is certainly correct in stating that: "the ability to distinguish between phonemes ... does not in itself imply an ability to understand verbal messages". On the other hand, it is felt that this short test of gross discrimination will form a useful input to the overall student profile.

SYNONYMOUS SENTENCES (4B)

45. Nevertheless, we agree also with Harris (1969) that: "Sound discrimination tests are very useful on the beginning and intermediate levels of instruction .... on the higher intermediate and advanced levels we are more concerned with measuring general comprehension of complete utterances and sequences of utterances." Hence we come to the first of three more general tests of listening comprehension. This is a test based on synonymous sentences and it must immediately be acknowledged that this is an impure test, depending as much on reading ability as on listening. If one wishes to conduct this type of test, however, there is no acceptable alternative form. Pictures bring their own cultural and practical difficulties and aural distractors would impose too great a strain on short term memory. And again there are counterarguments against the objection to impurity. The students are at an intermediate level and should have little difficulty with the lexical content of the options. More important, a test of an integrated skill is, by its very nature, a relatively indelicate test unless its content is trivial. It therefore seems unrealistic to insist on separating one linguistic skill from others as if each were in separate and watertight compartments, instead of being different realisations of the same underlying linguistic competence.

46. The test consists of 10 items, requiring 30 seconds each, a total of 7 minutes including instructions and examples. Several criteria were used in the construction of the test. First the test is intended to measure general auditory comprehension and so its content avoids lexical difficulty and does not focus on any particular linguistic elements. Secondly, since this is an aural test, the printed options are kept structurally simple. Any difficulty lies in the unravelling of semantic content from the structure of the aural stimulus. Thirdly, Rivers (1968) warns that short-term memory in a foreign language is much shorter than in L1; for this reason care has been taken not to impose any memory strain by making the stem too long. The longest stem is 13 words and the mean for the 10 items is 10.0.

APPROPRIATE RESPONSES (4C)

47. The second test of the general listening skill is Test 4C, Appropriate Responses, which in format is very similar to the previous test. It possesses also the same impurity in that it requires the ability to select from four printed responses. The arguments used above apply equally here and will not be repeated except to stress

that the reading requirement has been kept to a minimum by making the options lexically and structurally simple. Further, ample time is allowed for the students to read them. Speaking of this kind of test, Harris (1969) says: "there is evidence that even a short test of this nature is both a reliable and valid measure of general comprehension."

48. The test comprises 10 items and requires a total of 7 minutes to complete. It contains yes/no and a variety of Wh- question forms, and the 10 items make up a more or less continuous conversation.

LISTENING COMPREHENSION - SPATIAL TERMS (4D)

49. This is the last of the three tests of general listening comprehension, and it has two distinctive features. First it is a much more pure form of test than the previous two; the first part requires non-linguistic responses for which very little reading is needed, and the second part is multiple choice but again with minimal reading requirement. The other difference is that, although this is a test of general listening ability, it also focusses very specifically on one linguistic element - directional and other spatial terms, which are of prime importance in air traffic control. It does not, however, depend entirely on knowledge of compass directions. Eleven of the eighteen questions demand this knowledge, seven do not. Of these seven, five require identification of features in one form or another, and two require simple measurements. Lexically and structurally the test has been kept simple because it does contain a great deal of listening. The map contains some redundant features intended to aid comprehension, and others to act as distractors. To avoid visual confusion, however, these have been limited.

50. The test requires a total of 10 minutes to complete (3 minutes for the first part, 4 for the second and two sets of instructions needing 1½ minutes each). It also requires two distinct voices on the tape one to give the instructions, the other to read the exercise. In practice the test will doubtless need considerable amendment, but it appears to meet the particular needs of the students and could be a useful test both of spatial terms and of the listening skill.

NUMERAL TRANSCRIPTION (TEST 4E)

51. We leave the general tests and go on to the first of two highly specific tests of selected linguistic elements. Test 4E focusses on a feature of paramount importance to controller - an accurate and speedy facility with numerals. There are 30 items in the test, the first 20 requiring the transcription into figures of integers, starting with simple 2 digit items and working up to 5 figure items. This covers the whole range of numbers commonly used in aviation in contexts such as communication channels, circuit information, fuel states, altitudes and distances. The second part of the test requires the transcription of simple vulgar and decimal fractions - fractions

because they feature for example in meteorological information such as cloud cover, decimals because they are used inter alia for radio frequencies.

52. An aural test, with a linguistic auditory input and a non-linguistic response, seems to offer the simplest and most effective test of a student's knowledge of numerals and a transcription test avoids the temptation of testing numeracy. It is presumed that in the great majority of cases a receptive test of this particular kind will reflect also the student's productive ability. The test requires 7 minutes to administer, including examples and instructions.

TIME (TEST 4F)

53. The final test focuses also on a specific linguistic element - time - but to some extent this tests the general skill as well. It will be essential for the controller to have a complete mastery of the 24 hour clock in his duties. This test is designed to find out whether the student is proficient in this and the other two forms of telling the time commonly used in English. An apparently simple way of doing this would be to give the student a series of blank clock faces and then read out times to him, instructing him to add the hands of the clock in the appropriate places. In practice this creates difficulties which would not occur, for instance, in setting the hands of an actual clock. Thus, given a time such as 6.45, there is a danger in a two dimensional representation, of drawing the hour hand pointing to 6 before drawing the minute hand pointing to 9. The result is ambiguous between 5.45 and 6.45.

54. For this reason the method used in the test was preferred. Some numerical ability is clearly called for by this method, but this has been minimised by using only simple round figures in the time differentials. Furthermore, the structural and lexical content are absolutely minimal and each item is repeated after a short pause. The test requires 7 minutes to complete, including instructions and examples.

SECTION V: CONCLUSION

55. In conclusion we consider the test battery again from a broad perspective and draw together some of the main points made in this paper. The first point to re-emphasise is the integration of testing in the whole training programme. Heaton (1975) begins his book by stating that: "Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other." This can be a factor either for good or bad. In some cases the backwash effect by which the content and even the teaching methods of a particular course are determined by the final examination can exert a misdirected or disproportionate influence. If course objectives are to be determined by the target examination it is essential that that examination provides relevant and meaningful goals for the students. In other cases, however, course objectives can be determined by assessing the students' need in other ways. This is so with all courses in ESP, where the training specification can be

constructed on the basis of skills and linguistic needs analysis. Here the testing element of the course should be the servant of teaching.

56. In the case of English for Air Traffic Control, the syllabus is based entirely on needs analysis, and is continually amended to reflect changing needs. Testing therefore exerts virtually no back-wash effect. The main aim of the initial test battery is diagnostic and the results will provide the first major input to individual student profiles. Thus throughout the course it should be possible to compare a student's profile, representing his current state, with an air traffic controller's profile representing a linguistically adequate controller.

57. Because of the fundamental role the initial test battery is designed to play, it requires a considerable amount of time to complete. Adequate time was regarded as a necessary, though not a sufficient condition. As Heaton says: "The longer the test, the more reliable a measuring instrument it will be - although length itself is no guarantee of a good test." The tests require a total of 3 hours 50 minutes but are spread over at least three days. It might be noted that the times allotted to the four tests do not correspond in proportion with the importance weightings allocated to the four linguistic skills. One reason for this is that it was considered desirable to get a reasonable balance between subjective and objective assessment and some aspects tested do not lend themselves readily to objective testing. This is one form of external constraint on the structure of the tests. Another reason is that the tests are aimed to assess both language skills and linguistic competence - that is to say knowledge of the elements of the language. Some elements are more easily tested, for example, by means of a reading test than by other means. Thus the fact that the tests of reading comprehension take 65 minutes, although the reading skill was given only a 20% importance rating, reflects this factor.

58. The choice of structure and content of the test battery was made on the basis of the assessed needs of the students. Lado (1961) says: "The decision between testing skills or testing elements, if a decision has to be made, will be made on the basis of what information we need concerning the student's knowledge of the goal language." In this case the decision was made not as to the choice between skills and elements, but as to the proportions of each, and it was made not only on the basis of the students' knowledge but also on the basis of his needs. (The fact that at this stage these assessments were made at least partly subjectively is not important.) Harris (1969) says: "... great care must be taken to ensure that the final form of the test includes a broad range of relevant grammatical problems in proportions which reflect their relative importance." (author's underlining). This test battery attempts to do this.

59. Some thought was given to the means of presenting the listening comprehension tests. In its present form it is presented entirely by means of tape recorder. Valette (1967) makes the point that this

form of presentation of aural material makes comprehension more difficult since the student loses the non-linguistic clues that are part of normal conversation. Nevertheless, tape recording allows an administratively simpler and more consistent presentation of tests than does personal presentation. It was mentioned earlier, however, that the school possesses video facilities, including closed circuit television, and it is possible that parts of the battery could usefully be presented by this means in practice.

60. Finally, it should be repeated that the test battery in its present form represents only a first draft, and in validating it in practice it will doubtless be necessary to make a great number of amendments. Nevertheless, it is believed that the role of testing on this particular course has been correctly assessed and the method of test construction offers a rational and systematic approach to the problem.

References

Davies, A. (1964) English proficiency test battery (EPTB)

Harris, D.P. (1969) Testing English as a second language. McGraw-Hill.

Heaton, J.B. (1975) Writing English language tests. Longman.

Ingram, Catford, Mackin & May (n.d.) English language battery (ELBA)

Lado, R. (1961) Language Testing. Longmans.

Lee, W.R. (1957) The linguistic content of language teaching. ELT 1957 II.

Rivers, W. (1968) Teaching foreign languages. University of Chicago Press.

APPENDIX 'A'

PART I: Transcript of tests 4A - 4F.

PART II: Sample of Student's test booklet

Appendix I

INSTRUCTIONS FOR TEST 4A

This is Test 4A. Look at example A in the box at the top of page 1. Example A. 3 words are printed there. Listen carefully as I read one of those words. BARK (pause 4 secs) PARK. The word I read was PARK .... P-A-R-K ... the second word. So that word is underlined. Now look at Example B ... Example B. Listen carefully as I speak one of the three words. Draw a line under the word you hear. MINE .... 4 secs ... MINE. The word I read was MINE, the first word. So you should have drawn a line under the first word

We shall now go through items 1 to 40. For each item I shall read one word twice. Listen carefully and draw a line under the word you hear. Ready?

1. PEAT ... PEAT
2. BOAT ... BOAT
3. LOCK ... LOCK
4. SHOE ... SHOE
5. AT ... AT
6. PILE ... PILE
7. MEET ... MEET
8. HEART ... HEART
9. PAIR ... PAIR
10. CART ... CART
11. FIGHT ... FIGHT
12. CALL.... CALL
13. BULL ... BULL
14. PAINT ... PAINT
15. FEAR ... FEAR
16. COOT ... COOT
17. DEAN ... DEAN
18. OUR ... OUR
19. BUILD ... BUILD
20. FIRST ... FIRST
21. FRY ... FRY
22. PIT ... PIT
23. DRAIN ... DRAIN
24. GRAPH ... GRAPH
25. CARTER ... CARTER
26. CHEST ... CHEST
27. CLAP ... CLAP
28. DIN ... DIN
29 BRINK ... BRINK
30. BAND ... BAND
31. FATE ... FATE
32. BID ... BID
33. BRUTE ... BRUTE
34. PRICE ... PRICE
35. BUTTER ... BUTTER
36. GAIN ... GAIN
37. TIN ... TIN
38. BATH ... BATH
39. CLOWN ... CLOWN
40 LATTER ... LATTER

That is the end of Test 4A. Turn now to page 2 for Test 4B. Turn to page 2.

(TRANSCRIPT OF TEST 4B)

This is Test 4B. Look at example A in the box at the top of page 2. In this test I shall read a short sentence. Listen carefully to the sentence and then choose which one of the 4 sentences printed in your book has the same meaning as the sentence you hear. Which sentence means the same. Listen now to the sentence for example A:

I WISH I HAD COME TO THE CINEMA WITH YOU LAST NIGHT.

The correct choice is sentence c so a cross has been put in the brackets next to sentence c.

Look at example B and listen carefully while I read the sentence:

IF JOHN HAD STAYED LONGER HE WOULD HAVE MET BILL.

Put a cross next to the sentence with the same meaning. 20 secs .....
The correct answer was sentence d: JOHN LEFT BEFORE BILL ARRIVED so you should have put a cross next to sentence d.

We shall now go through items 1 to 10. Listen carefully - each sentence will be read only once. Put a cross next to the sentence you choose.

1. IT'S 8 O'CLOCK ALREADY. JOHN SHOULD BE HERE BY NOW.

25 secs

2. WE WOULD SWIM MORE IF THE WEATHER WERE WARMER.

25 secs

3. WE INTENDED GOING TO THE CINEMA BUT DAVID WAS TOO TIRED.

25 secs.

4. IS JOHN IN THE AIR FORCE AS WELL AS DAVID?

25 secs.

5. IF YOU HAD TOLD ME I WOULD HAVE HELPED YOU.

25 secs.

6. DAVID IS GOING TO BE TALLER THAN HIS FATHER.

25 secs.

7. JOHN DOESN'T PLAY THE PIANO AS WELL AS DAVID DOES.

25 secs.

8. IF I HAVE TIME I HOPE TO VISIT ROME AS WELL AS PARIS.

25 secs.

9. APART FROM TENNIS AND CRICKET MY FAVOURITE SPORT IS FOOTBALL.

25 secs.

10. DAVID NEVER SMOKED UNTIL AFTER HE WAS MARRIED.

25 secs.

That is the end of test 4B. Turn now to page 4. Page 4.

50 50

This is Test 4C. In this test there is a conversation between two men. You will hear the first man speak. You must then choose the best reply for the second man from the 4 sentences printed in your book.

(Repeat)

Look at example A in the box at the top of page 4. Example A. Listen carefully to the first man's sentence and then choose the best reply for the second man:

DAVID, HOW NICE TO SEE YOU! HOW LONG HAVE YOU BEEN IN NOTTINGHAM?

The best reply is ALMOST A YEAR so sentence b has a cross next to it.

Now look at example B and listen carefully to the first man's sentence. Then choose the best reply and put a cross next to it:

WHERE DID WE LAST MEET?

The best reply is sentence d: IN BIRMINGHAM. So you should have put a cross next to sentence d.

Now we go on to items 1 to 10. Listen carefully to each sentence and put a cross next to the best reply. Each sentence will be read only once.

1. I'VE JUST HAD LUNCH. HAVE YOU HAD YOURS YET?
2. WOULD YOU LIKE TO HAVE ANOTHER CUP OF COFFEE?
3. WHY ARE YOU IN SUCH A GREAT HURRY?
4. WHICH TRAIN DO YOU HAVE TO CATCH?
5. HOW OFTEN DO YOU HAVE TO GO TO LONDON BY TRAIN?
6. DO YOU MIND HAVING TO TRAVEL BY TRAIN?
7. I HAVEN'T SEEN YOUR SISTER FOR YEARS. WHERE IS SHE NOW?
8. REALLY? WHAT'S SHE DOING THERE?
9. HOW OFTEN DO YOU HAVE LUNCH HERE?
10. WHY DON'T YOU PHONE ME WHEN YOU GET BACK FROM LONDON?

That is the end of Test 4C. Turn now to page 6.

(Transcript of test 1D)

This is test 4D. Look first at the map on page 6 - page 6. In the first part of this test you have to mark certain points on the map. Listen to Example A:

LOOK AT CAT ISLAND. ON THE SOUTH COAST THERE IS A VILLAGE. PUT A FIGURE 1 NEXT TO THIS VILLAGE.

You see that a figure 1 has been marked next to the village on the south coast of Cat Island. Listen next to Example B:

ON CAT ISLAND THERE ARE SOME CLIFFS ON THE NORTH COAST. PUT A FIGURE 2 NEXT TO THESE CLIFFS.

You see a figure 2 next to the cliffs on the north coast of Cat Island.

Now listen carefully and follow the instructions in the rest of the exercise.

"Find Wolf Island. On the south east a river enters the sea. Mark this with a figure 3." (10 secs)

"On the south coast of Wolf Island there is a bay. Put a figure 4 in the bay" (10 secs)

"On Wolf Island there are two villages. Draw a line between these villages. Write a figure 5 on this line." (10 secs)

"Find Crab Island. On Crab Island there are 4 lakes. Mark the most northern lake with a figure 6". (10 secs)

"On the west coast of Crab Island is a bay. Put a figure 7 in this bay" (10 secs)

"Look at Fox Island. On the south west point there are some rocks. Mark these rocks with a figure 8." (10 secs)

"On Fox Island there is one village. Draw a line from this village to the nearest village on Crab Island. Mark this line with a figure 9." (10 secs)

"Halfway between Fox Island and Crab Island write a figure 10"

Now look at page 7 of your book. Here you have to answer some questions about the map. Find Example A on page 7. Listen to the question:

"How many villages are there on Cat Island?". The correct answer is 3. Answer b has been marked with a cross.

We shall now go through items 1 to 10. Listen carefully to the questions and put a cross next to the answer you choose.

1. Look at Cat Island. On which shore are the cliffs?
2. If you went from Wolf Island to Fox Island, in which direction would you be going?
3. How many rivers are there on Wolf Island?
4. How far is it from Crab Island to Fox Island?
5. Which is the smallest of the four islands?
6. There are four lakes on Crab Island. In which part of the island is the biggest lake?
7. How far is it between the two villages on Wolf Island?
8. On Cat Island there is one river. On which side of the island does it enter the sea?
9. Wolf Island has two bays. On which two coasts are they?
10. Look at Fox Island. What is marked on the north east point?

That is the end of test 4D. Turn now to page 8.

(Transcript for Test 4E)

This is test 4E. In this test you will hear some English numbers spoken and you have to write them down in figures. Look at Example A and listen carefully to the number:

TWO HUNDRED AND SEVENTY ONE

The number you heard was two hundred and seventy one so figures 2-7-1 have been written in the answer space. Now try Example B. The number will be spoken only once so listen carefully and write it down in figures.

FOUR HUNDRED AND FIFTY TWO

That was four hundred and fifty two and you should have written down figures 4-5-2. We shall now go through items 1 to 20. Listen carefully and write down each number as you hear it.

1. 43
2. 190
3. 238
4. 675
5. 403
6. 227
7. 951
8. 564
9. 386
10. 902
11. 1,260
12. 3,496
13. 2.354
14. 22,000
15. 34,000
16. 27,500
17. 4,650
18. 25,320
19. 14,527
20. 8,324

In the rest of this test instead of whole numbers you will hear fractions and decimal fractions. Look at Example C and listen carefully:

THREE TENTHS

You heard the fraction three tenths so it has been written as 3 over 10. Now you try Example D. Listen carefully to the fraction:

SEVEN EIGHTHS

The fraction you heard was seven eighths so you should have written down 7 over 8. 7 over 8.

Now look at Example E and listen:

THIRTY TWO POINT FOUR

The decimal fraction you heard was thirty two point four so it has been written down as 3-2-decimal point-4. Now you try Example F. Listen carefully:

TWENTY SEVEN POINT SIX

You should have written down twenty seven point six - that is 2-7-point-6.

Listen carefully now as we go through items 21 to 30:

21. three eighths. 22. four fifths. 23 two thirds.
24. five point seven. 25. two point eight five 26. forty nine point five
27. three quarters. 28. fifty five point seven five. 29. five eighths.
30. one hundred and two point nine.

THAT IS THE END OF TEST 4F. NOW TURN TO PAGE 9.

(Transcript for Test 4F)

This is Test 4F which tests your ability to tell the time in English. You will hear short messages given by aircraft pilots during flights. From these messages you will be able to work out the aircraft's take-off time, its arrival time, or the time of the message. (Repeat)

Look at Example A and listen carefully to the message. Each message will be read twice.

THE TIME IS NOW 3.15. I SHALL ARRIVE IN ONE HOUR EXACTLY. (5.secs)

THE TIME IS NOW 3.15. I SHALL ARRIVE IN ONE HOUR EXACTLY

His time of arrival will be 4.15. Answer d has therefore been marked with a cross. Now you try Example B. Listen to the message:

THE TIME IS QUARTER TO SEVEN. I TOOK OFF TWO HOURS AGO. (5.secs)

THE TIME IS QUARTER TO SEVEN. I TOOK OFF TWO HOURS AGO.

His take-off time was 4.45, so you should have put a cross next to answer c.

Now we shall go through items 1 to 10. Listen carefully to the messages. Each message will be repeated.

1. THE TIME IS NOW 9.30. I SHALL ARRIVE IN 3 HOURS. (5 secs). (Repeat)

(20 secs)

2. IT IS NOW 20 PAST FOUR. I SHALL ARRIVE IN ONE HOUR. (5 secs - repeat)

(20 secs)

3. IT IS NOW 11.35. I TOOK OFF TWO AND A HALF HOURS AGO. (5 secs... repeat)

(20 secs)

4. THE TIME IS NOW 1410. I'LL ARRIVE IN 4 HOURS. (5 secs...repeat)

(20 secs)

5. IT IS NOW 1745. I TOOK-OFF THREE AND A HALF HOURS AGO. (5 sec...repeat)

(20 secs)

6. IT'S EXACTLY 2230. I'LL BE THERE IN FOUR AND A QUARTER HOURS. (5 secs and repeat)

(20 secs)

7. IT'S JUST QUARTER TO EIGHT. I'VE BEEN FLYING FOR 2½ HOURS (5 sec.rep)

(20 secs)

8. I TOOK OFF AT 7.25 AND I'VE BEEN FLYING FOR 2 HOURS 15 MINUTES. (5 secs...repeat)

(20 secs)

9. I TOOK-OFF ONE AND A HALF HOURS AGO AT 8.40. (5 secs...repeat)

(20 secs)

10. IT'S 1920 AND I'VE BEEN FLYING FOR 2½ HOURS. (5 secs - repeat - 20 secs)

TEST 4A

EXAMPLES			
A.	bake	bark	back
B.	mine	nine	line

1.	pit	pat	peat	21.	fly	fry	pry
2.	boat	hoot	bait	22.	bit	pit	fit
3.	lock	lack	luck	23.	drain	grain	train
4.	show	shoe	shy	24.	grab	graph	grass
5.	at	art	ought	25.	charter	garter	carter
6.	pill	pile	pail	26.	jest	chest	zest
7.	mate	meet	mit	27.	clap	crap	trap
8.	height	hate	heart	28.	pin	din	tin
9.	pour	pure	pair	29.	brick	bring	brink
10.	cart	caught	cat	30.	bans	band	bank
11.	fit	fight	fate	31.	face	fade	fate
12.	cool	call	cowl	32.	big	bid	bit
13.	bell	bill	bull	33.	broom	brood	brute
14.	pant	pint	paint	34.	price	prize	prime
15.	fire	fear	fair	35.	butter	putter	mutter
16.	coat	cart	coot	36.	game	gain	gale
17.	den	dean	din	37.	thin	tin	din
18.	our	are	ore	38.	bath	bad	bat
19.	build	bald	bold	39.	frown	crown	clown
20.	fast	fist	first	40.	latter	ladder	lather

TEST 4B

EXAMPLE A.

a.	We all went to the cinema last night.	( ) a.
b.	I went to the cinema last night but you didn't.	( ) b.
c.	You went to the cinema last night but I didn't.	(X) c.
d.	No-one went to the cinema last night.	( ) d.

EXAMPLE B.

a.	John stayed longer than Bill.	( ) a.
b.	Bill stayed longer than John.	( ) b.
c.	Before John left he met Bill.	( ) c.
d.	John left before Bill arrived.	( ) d.

1.	a. John arrived at 8 o'clock.	( ) a.
	b. John has not yet arrived.	( ) b.
	c. John never arrives before 8 o'clock.	( ) c.
	d. John went home at 8 o'clock.	( ) d.
2.	a. We like to swim in warm weather.	( ) a.
	b. We often swim despite the weather.	( ) b.
	c. In future we are going to swim more.	( ) c.
	d. The weather will be good for swimming.	( ) d.
3.	a. We went to the cinema and David was tired.	( ) a.
	b. We plannnd to go to the cinema but did not go.	( ) b.
	c. David and I are tired of going to the cinema.	( ) c.
	d. It's very tiring going to the cinema.	( ) d.
4.	a. Both John and David are in the air force.	( ) a.
	b. John is in the air force but David isn't.	( ) b.
	c. David is in the air force but John isn't.	( ) c.
	d. David is in the air force and John may be.	( ) d.
5.	a. You told me but I couldn't help you.	( ) a.
	b. You didn't tell me so I didn't help you.	( ) b.
	c. I helped you because you told me.	( ) c.
	d. I helped you even though you didn't tell me.	( ) d.
6.	a. David is tall and so is his father.	( ) a
	b. David is tall but his father isn't.	( ) b.
	c. David is already taller than his father.	( ) c.
	d. David is not yet as tall as his father.	( ) d.
7.	a. David plays the piano but John doesn't.	( ) a.
	b. John plays the piano but David doesn't.	( ) b.
	c. John and David both play the piano.	( ) c.
	d. Neither John nor David plays the piano.	( ) d.

8. a. I am going to Paris and may go to Rome. ( ) a.
   b. I may go to Paris or Rome. ( ) b.
   c. I am going to Rome and may go to Paris. ( ) c.
   d. I am definitely going to Paris and Rome. ( ) d.

9. a. I like football more than tennis and cricket. ( ) a.
   b. Football is my favourite sport. ( ) b.
   c. Tennis and cricket are my favourite sports. ( ) c.
   d. I don't like tennis and cricket. ( ) d.

10. a. David is married and he smokes. ( ) a.
    b. David is married and he doesn't smoke. ( ) b.
    c. David is not married and he doesn't smoke. ( ) c.
    d. David smokes and is not married. ( ) d.

END OF TEST 4B

TEST 4C

EXAMPLE A.

a. Yes, I have.	( ) a.
b. Almost a year.	(X) b.
c. Until next September.	( ) c.
d. After the course ends.	( ) d.

EXAMPLE B.

a. From York to Birmingham.	( ) a.
b. 1973 I think.	( ) b.
c. Yes we did.	( ) c.
d. In Birmingham.	( ) d.

1.	a.	I'm pleased to hear it.	( ) a.
	b.	Yes I have, thank you.	( ) b.
	c.	That's kind of you.	( ) c.
	d.	I'm afraid I can't.	( ) d.
2.	a.	Yes, I think I have time.	( ) a.
	b.	Yes I have, thank you.	( ) b.
	c.	No, I prefer it black.	( ) c.
	d.	Not since breakfast.	( ) d.
3.	a.	Yes, I have to hurry.	( ) a.
	b.	I'm afraid it's too late.	( ) b.
	c.	I have to catch a train.	( ) c.
	d.	You're quite right.	( ) d.
4.	a.	The 2.30 to London.	( ) a.
	b.	Yes, I really must.	( ) b.
	c.	I have to change at Birmingham.	( ) c.
	d.	Because it doesn't stop here.	( ) d.
5.	a.	Just before Christmas.	( ) a.
	b.	I haven't got a car.	( ) b.
	c.	About once a month.	( ) c.
	d.	It's very expensive now.	( ) d.
6.	a.	Yes, I think of it a lot.	( ) a.
	b.	No, it's very convenient.	( ) b.
	c.	Yes, it is rather pleasant.	( ) c.
	d.	I'll try to remember.	( ) d.
7.	a.	I saw her only last month.	( ) a.
	b.	She's very well, thank you.	( ) b.
	c.	You may see her soon.	( ) c.
	d.	She lives in France.	( ) d.
8.	a.	She married a teacher in Paris.	( ) a.
	b.	Yes, she has lived there for two years.	( ) b.
	c.	She often tells me about it.	( ) c.
	d.	It's quite true, I agree.	( ) d.

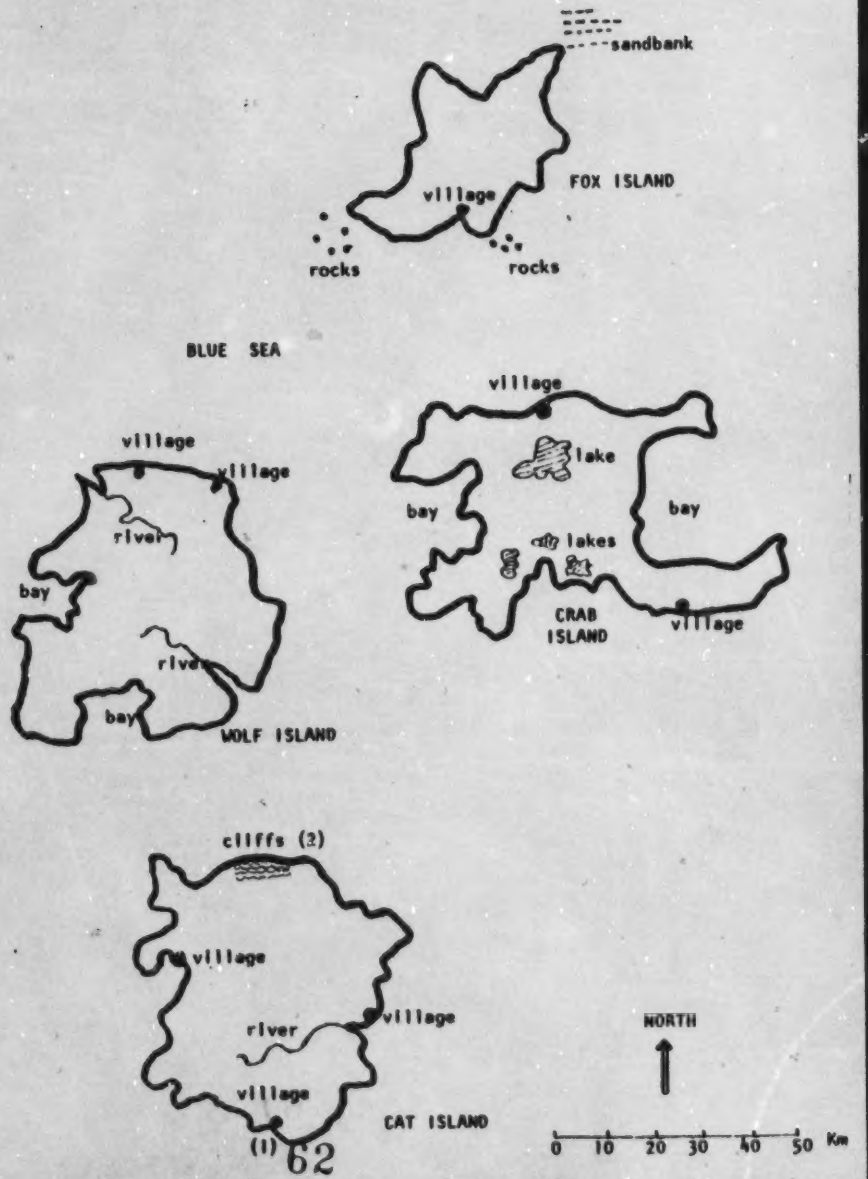
TEST 4C (continued)

9. a. No, it's too expensive. ( ) a.
   b. Yes, I try to when I can. ( ) b.
   c. They open for lunch at noon. ( ) c.
   d. Every Tuesday usually. ( ) d.

10. a. I'm sorry, I always seem to forget. ( ) a.
    b. Yes I will; give me your number. ( ) b.
    c. I don't know; I often think of it. ( ) c.
    d. Yes please do that and don't forget. ( ) d.

END OF TEST 4C

TEST 4D


sandbank
village
FOX ISLAND
rocks
rocks
BLUE SEA
village
village
river
bay
river
bay
WOLF ISLAND
village
lake
bay
bay
lakes
CRAB ISLAND
village
cliffs (2)
village
river
village
village
CAT ISLAND
NORTH
0 10 20 30 40 50 Km

TEST 4D PART II

EXAMPLE A

a.	2	( ) a.
b.	3	(x) b.
c.	4	( ) c.
d.	5	( ) d.

1.	a.	north	( ) a.
	b.	south	( ) b.
	c.	east	( ) c.
	d.	west	( ) d.
2.	a.	north east	( ) a.
	b.	south east	( ) b.
	c.	north west	( ) c.
	d.	south west	( ) d.
3.	a.	8	( ) a.
	b.	6	( ) b.
	c.	4	( ) c.
	d.	2	( ) d.
4.	a.	about 20 Km	( ) a.
	b.	about 50 Km	( ) b.
	c.	about 80 Km	( ) c.
	d.	about 120 Km	( ) d.
5.	a.	Cat Island	( ) a.
	b.	Wolf Island	( ) b.
	c.	Crab Island	( ) c.
	d.	Fox Island	( ) d.
6.	a.	south west	( ) a.
	b.	north	( ) b.
	c.	north west	( ) c.
	d.	south	( ) d.
7.	a.	5 Km	( ) a.
	b.	20 Km	( ) b.
	c.	50 Km	( ) c.
	d.	80 Km	( ) d.
8.	a.	north	( ) a.
	b.	south	( ) b.
	c.	east	( ) c.
	d.	west	( ) d.

9. a. north & west ( ) a.
   b. south & west ( ) b.
   c. north & east ( ) c.
   d. south & east ( ) d.

10. a. rocks ( ) a.
    b. sandbank ( ) b.
    c. wreck ( ) c.
    d. river ( ) d.

TEST 4E

EXAMPLE A

271

EXAMPLE B

---

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20 _____

EXAMPLE C.

$\frac{3}{10}$

EXAMPLE D.

EXAMPLE E.

32.4

EXAMPLE F.

21. _____
22. _____
23. _____
24. _____
25. _____
26. _____
27. _____
28. _____
29. _____
30. _____

END OF TEST 4E

TEST 4F

EXAMPLE A.		
Arrival time will be:	a. 3.30	( ) a.
	b. 4.30	( ) b.
	c. 4.00	( ) c.
	d. 4.15	(X) d.
EXAMPLE B.	a. 6.00	( ) a.
	b. 5.30	( ) b.
Take-off time was	c. 4.45	( ) c.
	d. 5.15	( ) d.

1.	Arrival time will be	a. 12.00	( ) a.
		b. 12.30	( ) b.
		c. 11.30	( ) c.
		d. 11.00	( ) d.
2.	Arrival time will be	a. 5.20	( ) a.
		b. 5.40	( ) b.
		c. 6.20	( ) c.
		d. 6.40	( ) d.
3.	Time of take-off was	a. 8.05	( ) a.
		b. 9.05	( ) b.
		c. 9.25	( ) c.
		d. 8.25	( ) d.
4.	Arrival time will be	a. 1414	( ) a.
		b. 1844	( ) b.
		c. 1610	( ) c.
		d. 1810	( ) d.
5.	Time of take-off was	a. 1700	( ) a.
		b. 1545	( ) b.
		c. 1445	( ) c.
		d. 1415	( ) d.
6.	Arrival time will be	a. 2245	( ) a.
		b. 0215	( ) b.
		c. 0245	( ) c.
		d. 0315	( ) d.
7.	The time now is	a. 5.15	( ) a.
		b. 4.15	( ) b.
		c. 4.30	( ) c.
		d. 5.30	( ) d.

8. The time now is

a. 8.40 ( ) a.
b. 8.50 ( ) b.
c. 9.50 ( ) c.
d. 9.40 ( ) d.

9. The time now is

a. 10.20 ( ) a.
b. 9.10 ( ) b.
c. 10.10 ( ) c.
d. 9.50 ( ) d.

10. Time of take-off was

a. 1500 ( ) a.
b. 1650 ( ) b.
c. 1950 ( ) c.
d. 2000 ( ) d.

Proposals for a proficiency test for use on pre-sessional courses

Lewis Karr

University of Essex Department of Language and Linguistics
Occasional Paper No.21 pp. 63-81

In a paper given at the 1976 AILA conference, Spolsky argues that modern language testing has developed through three very loosely defined stages: the pre-scientific, the psychometric-structuralist and the psycholinguistic-sociolinguistic. In the pre-scientific stage subjectivity held sway. The teacher was thought to require no special skills in assessing a student's language ability, other than those developed through experience in his teaching career, and on that basis he stood in judgement over his charges. The examination questions set were open-ended and usually included passages for translation, dictations and free compositions. The oral skills were rarely tested.

However, language teaching and testing methodology usually follows developments in relevant academic disciplines, e.g. linguistics, psychology and sociology, in some cases rather slowly. The emergence of the psychometric-structuralist stage is a reflection of the recognition that linguistic theory has some contribution to make to a theory of testing and that testing can be conducted in a scientifically exact manner. In the most recent developments in language teaching, the concern for scientific rigour has given way to a concern to assess a student's language proficiency in terms of his ability to communicate effectively in the language. In a sense the psycholinguistic-socio-linguistic stage is a synthesis of the two earlier stages, incorporating the structuralists' concern for objectivity and the relative freedom of expression encouraged in Spolsky's 'pre-scientific' stage.

The purpose of this paper is to design a test battery suitable for use as an initial entry test for a pre-sessional course for foreign graduate students. In sections 1 and 2, I will describe the students and the course, and in section 3, I will consider the type of test required. In section 4, I will discuss in detail the developmental stages in language testing mentioned above and evaluate the advantages and disadvantages of discrete item tests and integrative skills tests. In the final section, I will present examples of the testing devices which I would wish to include in a proficiency test for adult students.

1.0 The Students

1.1 The students for whom the test is designed are all British Council Study Fellows who will be following a twelve week language improvement course prior to taking up places on post-graduate courses in the social sciences at British universities. They will come from various educational and cultural backgrounds and from countries in South and Central America, South East Asia and the Middle East. The languages represented will therefore be Spanish, Arabic, Farsi, Thai, Indonesian and Japanese.

1.2 Previous learning experience

All of the students are graduates from universities in their own countries, or from universities in other third world countries. The majority will have studied English as a service language during their undergraduate careers, but few will have been examined in English in their finals. They will also have studied English for some years in secondary schools and most will have taken English classes at either British or American cultural centres in their own countries. The majority will have taken the EPTB (Davies Test) before they come to England. It can be assumed that they will have a reasonable knowledge of English, but that

they require an intensive course before they are ready to study in this country.

1.3 Language and language related needs.

Various surveyshave been carried out which seek to determine the language needs of foreign students studying in this country (e.g. White, Sinclair-Rogers). White's study concentrates on the specific skills required by students following different courses of study. He found that a post-graduate's initial requirement was for an adequate standard in listening and reading comprehension as most courses begin with an intensive lecture and reading programme. After the initial framework of the course has been defined the student requires the productive skills necessary for participating in seminars and producing written assignments.

In the second study Sinclair-Rogers investigated the linguistic needs of foreign students in both social and study situations. He found that most foreign students experience difficulty in fluent communication at an informal level, in formal subject discussion and seminars and in the expression of personal viewpoints. They are also likely to limit themselves to monosyllabic answers where more informative and extended responses are required, to have little knowledge of topics of interest commonly discussed by educated speakers and to have difficulty with various conventions of non-verbal communication.

2.0 The Course

The course lasts for twelve weeks and is divided into three four-week blocks. In the first block the students spend the first week following a social survival course in which they are introduced to the various services available to them, e.g. the banks, public transport, shops, etc., and the language associated with them. In the remaining three weeks of the first block they receive further practice in social English and social conventions are discussed.

The second block is given over to language development with particular reference to the study skills they will require when following their academic courses, and to the language associated with their various fields of interest. In the final part of the course the components just described continue, but are much reduced. The students spend a large part of this section doing background reading in their subject in preparation for a 4,000 word project which is handed in at the end of the course.

During the first block the students are assigned to classes on the basis of an entry test. In the middle and final sections of the course they are regrouped according to their academic interests. Where possible, i.e. when there are enough students studying one subject, the students are also graded according to their proficiency in English, but this is often not possible to achieve.

As noted in the introduction, the purpose of this paper is to discuss and construct a suitable entry test for the course just described. It is to this that I will now turn.

3.0 What sort of test?

Language tests fall into two main categories: predictive tests and achievement tests. The first category is usually realised by an aptitude test such as the MLAT (Carroll and Sapon) or LAB (Pimsleur) which predict a student's likely success or lack of success in learning a foreign language. Achievement tests can be subcategorized into attainment, diagnostic and proficiency tests. The domain of an attainment test is determined by the course of study the students have followed and provides information concerning their success or failure in relation to the course alone. A diagnostic test attempts to identify and isolate the areas in which a student is deficient in a language in order to provide an input for the teaching programme. The purpose of a proficiency test is to assess a student's ability in the foreign language, irrespective of any course he may previously have followed.

Given the information on the students' background, future needs and the nature of the course they are to follow, we can now consider which of the above tests will be suitable. An aptitude test does not seem to be what is required as the students have already spent a number of years studying English and because they have already been assessed as suitable candidates for further study in this country. Thus, the administration of a language aptitude test at this stage would be like slamming the door after the horse has bolted. As an initial test an attainment test is clearly out of the question as the students will not have followed any course in common before they arrive in England.

A diagnostic test is one often favoured in the situation described, as it provides specific information about the sort of errors the students make and indicates specific problems for the teacher to concentrate on. However, Upshur (1973) has argued that diagnostic testing is more properly regarded as a continuous process which takes place in all classroom interchanges between student and teacher and provides immediate inputs into the teaching. The point may best be clarified by a flow chart presented by Upshur which accurately captures the teaching situation. (See overleaf.) In this analysis the teacher and student constantly react with one another, the teacher providing an input for the student to work on, (the situation requiring behaviour), and the student in turn providing a response which acts as an input for the teacher, who analyzes it as correct or incorrect and presents further appropriate inputs for the student accordingly. It might be argued that a diagnostic test is better able to isolate the source of the student's problem accurately than is the teacher while he is in the process of teaching a class. This obviously depends on the teacher's competence in his art, but it also assumes that it is possible to isolate the causes of an error in a non-communicative situation (most diagnostic tests take the form of discrete item tests, - see below for discussion). This assumption seems to me to be highly dubious. The crucial decision points in Upshur's flow chart are those at which the teacher questions whether the problem is appropriate or whether the situation is appropriate. No matter how well constructed a diagnostic test is, it cannot take into consideration the appropriacy of the situation and the appropriacy of the problem set is, willy-nilly, subjective.

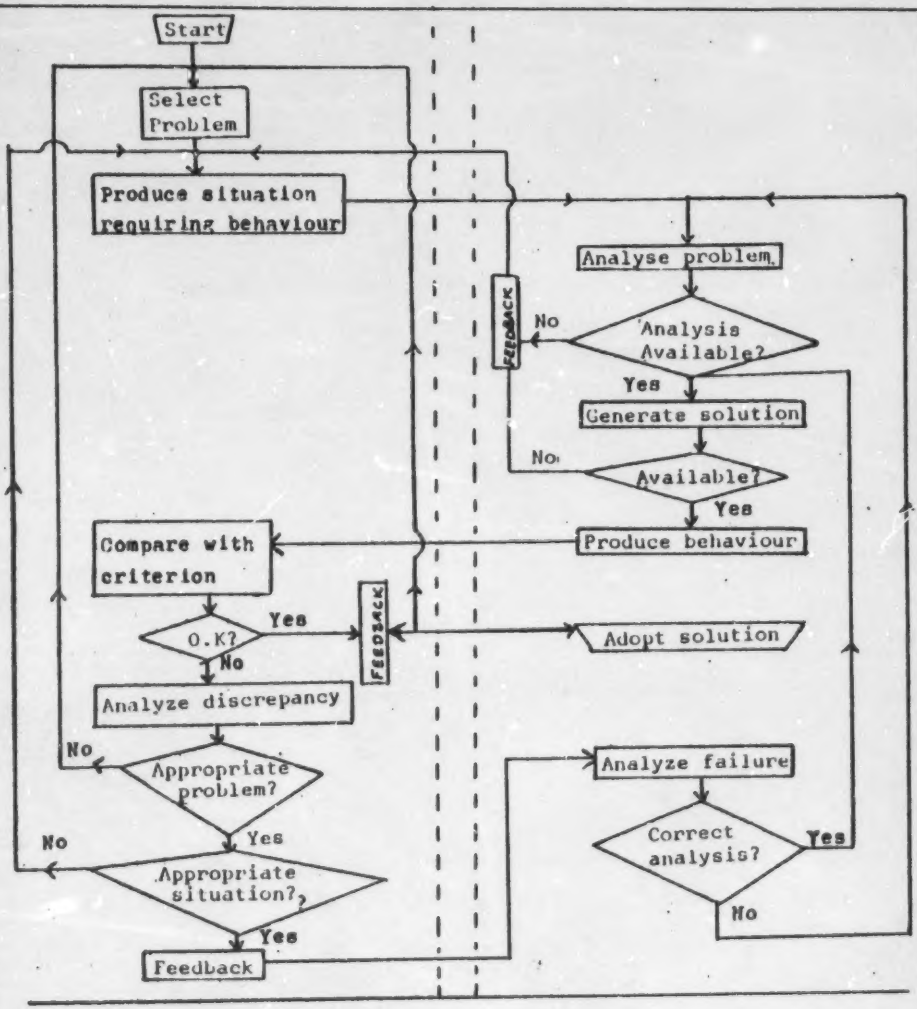


Table 1. (from Upshur, p.209)

Having rejected aptitude, attainment and diagnostic tests as being unsuitable for the purposes of the course described, we are left with the possibility of a proficiency test. It was noted in section 1.2 that most of the students on the course will have taken the EPTB before coming to England. One might question, therefore, the necessity of administering another proficiency test on their arrival in England. There are two reasons why I think it is necessary: firstly, some of the students will not have taken the test and those who have will inevitably have taken it in vastly different situations; secondly, for reasons that I will argue in detail below, the rationale behind the EPTB is now somewhat out of date and can be improved upon. The suitability of a proficiency test as an initial test for the course described is clear. As Heaton (1975,p.164) notes: "... the proficiency test looks forward, defining a student's language proficiency with reference to a particular task which he will be required to perform. Proficiency tests are in no way related to any syllabus or teaching programme." The tasks which the students will be required to perform have already been discussed. In the following section I will consider the type of proficiency test most relevant to the students' needs.

4.0 Proficiency Tests: - Discrete Item or Integrative Skills Tests?

One of Davies's criteria of adequacy for language tests is that they should have construct validity, i.e. that they should be dependent on a general theory of language and language learning. Examination of various tests (e.g. EPTB, ELBA) shows that the history of testing has largely followed the development of theories of language and psychology and associated theories of language teaching, albeit with something of a time lag. In this section I will consider the theoretical bases of discrete items tests and integrative skills tests, the rationale for each and their relative advantages and disadvantages. (By discrete item tests I mean those that test one item at a time, all other tests falling into the category of integrative skills tests.)

4.1 Discrete Item Tests.

The development of discrete item tests is usually credited to Fries and his sometime student, Lado. The linguistic theory upon which discrete item tests are based is structuralist linguistics of the school of Bloomfield, and the psychological theory is that of behaviourism which found its most extreme expression in Skinner's 'Verbal Behaviour' (1957). Briefly, structuralist linguistics analyzes the language as a hierarchical system with the sentence at the highest level and the phoneme at the lowest. The successful learning of a language requires the mastery of the systematic relationships between the constituents of the sentence. Behaviourist psychology argues that learning is a process of habit formation in which the learner acquires automatic responses to stimuli provided by an interlocutor's speech, or the speaker's perception of his environment. In his book, Language Testing Lado (1961, p.22) makes reference to both these base theories:

> "The theory of foreign language testing is based on present linguistic understanding of language and on observations concerning the role of habit in learning a foreign language. ... (It) assumes that language is a system of habits of communication ... These habits involve matters of form, meaning and distribution at several levels of structure, namely those of the sentence, clause, phrase, word, morpheme, and phoneme." (p.22)

In his discussion of what it means to 'know' a language Fries (1945,p.3) argues:

> "A person has 'learned' a language when he has thus first, within a limited vocabulary, mastered the sound system, (that is, when he can understand the stream of speech and achieve an understandable production of it) and has, second, made the structural devices (that is, the basic arrangement of utterances) matters of automatic habit."

The theoretical implications of the contemporary linguists' and psychologists' view of language found expression in both the teaching methodology and the testing of language proficiency in the 'sixties. The theory of habit formation led to an interest in making contrastive linguistic analyses of languages in order to identify the respects in which they differed from one another and the respects in which they were similar. Lado notes that, "the complex process of communication through language as we know it is made possible by the system which operates largely without our awareness."(op.cit.p.13) This leads him to argue that in learning a second language the student will transfer 'the system of habits' he has acquired in his first language to the perception and production of the second language. Thus, to the extent that his native language differs from the target language the learner will encounter difficulties, while in the cases in which the learner's first language uses similar systems to those in the second language his difficulties will be correspondingly reduced. In language teaching this led to an attempt to isolate and concentrate on the areas of potential difficulty for the learner, and in language testing these areas were again identified as those which would best reveal the learner's proficiency in the second language.

The hierarchical analysis of language presented by structuralist linguistics led directly to the teaching and testing of discrete items. Again, Lado expresses the implications of linguistic theory to the applied art: "Utterances can transmit an almost infinite number of meanings and combinations of meanings, yet all the meanings conveyed through language in any utterance are transmitted through a limited number of different sounds (or written symbols). The minimum sound unit that can change one utterance into another in a language is called in linguistics a phoneme."(ibidp.6) Lado goes on to enumerate the other elements in the hierarchy which can change one utterance into another: morphemes, phrases, clauses, sentences and intonation. The best and most accurate method of assessing whether a foreign learner has acquired an unconscious awareness of the system of the target language, it was argued, is to isolate the different levels in the system and to test each independently as a discrete item. Examples of such items can be found in many of the standard texts on testing, e.g. Lado, Harris, Valette, Heaton. The following are from Harris.

1. Auditory discrimination test

The testee hears three words, e.g. 'cot' 'caught' 'cot', and his task is to determine whether one differs from the others, and if so, to indicate which one is different.

2. Grammatical structure.

Mary _____ in New York since 1960.

a) is living

b) has lived

c) lives

3. Sentence interpretation

An old friend of John's brought <u>him</u> news of his uncle

<u>Him</u> refers to a) an old friend;

b) John;

c) the uncle.

4. Word order.

When _____ _____ _____ _____?

a) plan; b) you; c) to go; d) do

5. Vocabulary

A brief, light sleep

a) yawn; b) nap; c) struggle; d) hug

6. Subject-verb agreement

The design of the two new bridges (are/is) very unusual.

4.1.1. The advantages of discrete item tests.

If one accepts the linguistic theory upon which discrete item tests are based, then their primary advantage lies in the fact that the item tested is pure, i.e. one and only one item is tested at a time. Thus, the testee's proficiency or deficiency can be isolated for a single item. However, as was noted in the discussion on diagnostic tests, while it may be possible to isolate an error, a discrete item test can give no indication as to the cause of the error. The other advantages of discrete item tests stem largely from the first. They are very easy to administer to a large number of testees at one time and, as there can be only one correct answer, little time is needed for marking and different markers can be relied upon to give the same score to any one testee. Moreover, a number of investigators have shown that the different sub-tests in a discrete item battery correlate highly with each other, thus indicating that such a test has overall reliability.

4.1.2 The disadvantages of discrete item tests.

The fundamental disadvantage of discrete item tests, and the reason why they will be rejected in the battery proposed at the end of this paper,

is that they fail to reflect actual language use. By definition, discrete items are divorced from the communicative use of language, demanding that the student fulfil highly artificial tasks in which the strategies he adopts in completing them are uncontrolled, and to a large degree, unknown. To make generalizations about a student's proficiency in language as a communicative tool from the result he achieves in a discrete item battery seems very dubious. In common with all multiple choice items, discrete items have the disadvantage that they demand a high degree of skill in their construction. To ensure that the correct answer is the one selected by good students and that the distractors adequately fulfil their function, each item must be proofed and pre-tested, which requires considerable expenditure of time and money. Where the population to be tested is large such expenditure may be justified, but for the situation that I have in mind, it is not. However, this is a practical consideration. As I noted above, the most serious problem regarding discrete item tests is a theoretical one. Both the linguistic and psychological foundations of such tests have been superceded. It is to the consideration of more modern theories that I now turn.

4.2 Integrative skills tests.

The Chomskyan revolution which gathered momentum in the sixties led to a reappraisal of both the linguistic analysis of language and, more importantly from the point of view of testing, the psychology of language learning. Chomsky argued (1957, 1965) that language is generative and characteristically novel. That is to say, he refutes directly the Skinnerian view that language learning is a habit-forming process. Rather he argues that language is a rule-governed system which, once acquired, provides the speaker with the means of constructing an infinite set of novel utterances. In his analysis of language Chomsky sets up a dichotomy between competence and performance. As these concepts have been explicitly adopted by at least one applied linguist in building a theory of language testing, it is worth quoting Chomsky at some length on the distinction he makes between them. In 'The Sound Pattern of English' (1968,p3) he notes:

> "The speaker produces a signal with a certain intended meaning; the hearer receives a signal and attempts to determine what was said and what was intended. The performance of the speaker or hearer is a complex matter that involves many factors. One fundamental factor involved in the speaker/hearer's performance is his knowledge of the grammar that determines an intrinsic connection of sound and meaning for each sentence. We refer to this knowledge - for the most part, obviously, unconscious knowledge - as the speaker/hearer's 'competence'. Competence, in this sense, is not to be confused with performance. Performance, that is, what the speaker/hearer actually does, is based not only on his knowledge of the language, but on many other factors as well, factors such as memory restrictions, inattention, distraction, non-linguistic knowledge and beliefs and so on. We may, if we like, think of the study of competence as the study of potential performance of an idealized speaker-hearer who is unaffected by such grammatically irrelevant factors."

Applying the implications of competence as the unconscious knowledge of a system of rules to testing, Spolsky (1968) relates the concept to the phenomenon of redundancy in language. He notes that, as many investigators have shown, the surface structure of an utterance incorporates a great deal of redundancy, the purpose of which is to protect a message when it is most vulnerable, i.e. in the actual transmission of the message from speaker to hearer. Moreover, he notes that when the redundancy in a language is reduced, by adding noise to a spoken text or by randomly masking words in a written text, a native speaker of the language has little difficulty in reconstructing the original text. This, Spolsky argues, is possible because of the native speaker's competence. His competence provides the native speaker with what Oller has described as a grammar of expectancy. It follows from this that a foreign learner's proficiency in a second language can be assessed by presenting him with a reconstruction task.

Tasks of this sort will be described and discussed below. Before moving on to this, however, a few points should be made concerning Spolsky's adoption of Chomsky's interpretation of competence. As we saw in the passage quoted above, Chomsky excludes from the domain of competence such 'grammatically irrelevant factors' as memory restrictions, nonlinguistic knowledge and beliefs. This delimitation of competence may be necessary on theoretical grounds, but empirical evidence suggests that if the competence model is to be adopted as an explanation of the native speakers' grammar of expectancy, its domain will have to be extended. The need for this can be demonstrated by considering the strategies employed in completing the following extract from a cloze test:

> Taking the birth rate first, Malthus __(1)__ that when incomes rose, people could __(2)__ to marry earlier. The effect of __(3)__ would be that more of a __(4)__ childbearing years would be spent in __(5)__ and, therefore, she would be exposed __(6)__ the risk of having children .....

The information necessary for supplying the word in (1) is largely syntactic, the word must be a verb, but if the correct verb is to be selected the testee will require some understanding of the notion of appropriacy. For example the verb 'hope' would satisfy the syntactic requirements of the sentence, it can be followed by a that clause, but it would be quite inappropriate in the context. The concept of appropriacy requires that the verb should be selected from a small group of verbs: think, believe, say, argue, demonstrate, etc. The selection of the correct word, in this case 'argue', requires more delicacy in the handling of appropriacy for the testee must appreciate the register of the text. Finally, the selection of the correct tense requires some knowledge of the world, i.e. the knowledge that Malthus is dead and therefore, that the past tense would be most appropriate. Similarly, in (2) the testee requires a degree of grammatical competence, but also an appreciation of the concept of lexical cohesion, realized in this case by collocation. The word missing is 'afford' which can be predicted because of the cohesive tie it forms with 'incomes rose'.

the selection of the correct word for (3) also involves an appreciation of cohesion as the required word makes anaphoric reference to the preceding sentence. (4) also requires some knowledge of the world, i.e. it is only women who bear children. The selection of the correct word in (5) also requires an understanding of cohesive relationships, this again is an example of collocation, the tie being made with 'marry' in the preceding sentence. In addition, the testee must have some knowledge of societal institutions. It is only in the case of (6) that grammatically formal information alone is required. The word missing is a form word, 'to'.

Further evidence of the need to extend Chomsky's notion of competence when relating it to integrative skills tests is provided by Upshur (1969). In administering a test of productive communication to a population of foreign learners of English he found that, "students with only a limited command of English could communicate perfectly if they were given enough time. Individual differences in amount of communication began to appear as progressively more severe time constraints were imposed." This would suggest that memory restrictions and speed of processing verbal information play some part in the competence of the speaker.

On theoretical grounds Hymes (1972) has argued that Chomsky's notion of competence is an inadequate description of the rules underlying actual communication. He proposes that: "... if linguistic theory is to be integrated with a theory of communication and culture .... four questions arise:

1. Whether (and to what degree) something is formally possible;
2. Whether (and to what degree) something is feasible in virtue of the means of implementation available;
3. Whether (and to what degree) something is appropriate (adequate, happy, successful) in relation to a context in which it is used and evaluated;
4. Whether (and to what degree) something is in fact done, actually performed, and what its doing entails?"

This expanded view of competence, (Hymes refers to it as 'communicative competence'), incorporates many of those factors which Chomsky relegates to performance. However, I would argue that it is just those considerations laid out in Hymes' points 1 - 4 that enable a native speaker to decode the communicative function of an utterance in many and varied contexts.

However, Chomsky's assertion that language is a rule-bound system which enables the speaker/hearer to generate and comprehend an infinite set of novel utterances remains relevant to the theory of a grammar of expectancy, which in turn explains the native speaker's ability to reconstruct messages in which redundancy has been reduced. In terms of testing this has led to the adoption of a number of devices which seek to tap the testee's underlying communicative competence, notable among them being the oral interview, variants of the cloze procedure,

dictation and composition writing. All of these are designated integrative skills tests for the obvious reason that each tests a number of skills simultaneously. I will consider each in turn.

The most straightforward means of assessing a student's communicative competence is, as Spolsky (1973) notes, to give the student an oral interview which requires the natural functioning of language. This method has not proved very popular because of the difficulty involved in administering such a test to a large population of testees. However, the latter problem does not arise when the number to be tested is relatively small and, in recent years the former problem has been attacked with the development of objective means of assessing the testee's performance. Early developments have been reported by Upshur (1969). The basic technique he employed was to present the testee with four pictures which differed in certain respects. The examinee's task is to give a single sentence description of one of the pictures, selected randomly, to the examiner who does not know which picture has been selected. On the basis of the testee's description the examiner decides which picture is being described, and if he selects the same picture as the one the testee was directed to describe the testee is scored accordingly. Upshur notes that in such a test the scorer reliability is satisfactory (0.99). He also notes that when correlated with scores achieved by students taking a discrete item proficiency test, those students who show a wide range of proficiency according to the discrete item test are capable of effective communication in a short time. Those students who achieve lower scores on a discrete item test can communicate as effectively as the others as long as time constraints are not imposed.

A more natural but administratively more complex interview technique has been developed by Walters (1978). In this an independent assessor is required in addition to the interviewer. Walters has analysed the communicative possibilities available to the testee, e.g. responding to the interviewer's directions or questions either correctly or incorrectly or appropriately or inappropriately, or by asking for clarification or repetition, or initiating an interchange with the interviewer. The assessor records the interlocutions between the examiner and examinee throughout the interview at three second intervals by noting down numbers which relate to the possible communicative acts performed by the participants. The proposed coding system is as follows:

1. Student responding. Correct.
2. Student initiating. Correct.
3. Student responding. Incorrect.
4. Student initiating. Incorrect.
5. Student initiating. Inappropriate.
6. Student responding. Inappropriate.
7. Student - Incomprehensible.
8. Interviewer accepting.
9. Interviewer questioning.
10. Interviewer repeating.
11. Silence.

At the end of the interview the numbers are plotted onto a matrix which gives a representation of the interview in terms of the extent of the testee's contribution and his ability to communicate correctly and appropriately, and the extent of the interviewer's contribution to the interview.

Obviously the administration of such a test of productive communication requires considerable time and expense in the training of the assessors. It is also inefficient in that it requires an examiner/examinee ratio of 2:1. However, when the population to be tested is small these problems do not seem to be insurmountable. Unfortunately there is no data available as yet regarding the reliability of a test of this nature, but it is hoped that the test will be administered to students joining the course described in section 2 above, and the necessary data obtained.

The cloze procedure has already been demonstrated briefly above. The procedure was originally developed by Taylor (1953) as a means of assessing the difficulty of written texts. However, it has been adopted by language testers for use as a test of both reading and listening comprehension. The basic technique in constructing a cloze test is to omit randomly every fifth, seventh, or ninth word in a written text. The student's task is to supply the missing word. There has been considerable debate as to how the test should be scored. The possibilities are that only the exact words restored to the text are counted as correct, or that any acceptable word is counted as correct. In the latter case the item has to be pre-tested with native speakers to ensure that an objective list of acceptable words is arrived at. Oller (1973) reports that in a study made at UCLA the best method was to accept any reasonable substitute for the missing word. He also found that a reading cloze test correlates most highly with tasks which require skill in listening comprehension.

The cloze procedure has been adopted as a device for testing listening comprehension skills by Templeton (1977). Administering the test to both native and non-native speakers at the University of Manchester, he found that the optimum number of words omitted from a spoken text is thirty and that they should be omitted at intervals of fifteen words. A greater density of omissions proved to be too taxing even for native speakers. The procedure which he found most effective was to allow the testees to hear the complete text once, read at normal speed, and then to replay the text broken down into sense groups with gaps of four seconds duration between each sense group. The words omitted in the text were replaced by 'bleeps'. He found that the results obtained correlated highly with teacher and peer ratings (0.91 and 0.85 respectively) and with the entrance test taken by foreign students studying at Manchester. Significantly the cloze procedure correlated most highly with the aural sub-test, (0.80).

The use of dictation as a vehicle for assessing a foreign learner's proficiency in English has recently found renewed popularity. Lado (1961)

dismisses dictation for the following reasons: " ..... it appears to measure very little of language. Since the order of the words is given by the examiner as he reads the material, it does not test word order. Since the words are given by the examiner, it does not test vocabulary. It hardly tests aural perception of the examiner's pronunciation, because the words can in many cases be identified by context if the student does not hear the sound correctly." (p.34) However, Oller (1972) has shown that it is a very useful device for testing overall proficiency, perhaps because one of the strategies adopted by the student is to gain as much as is possible from the context. On the question of the level of difficulty of the text, Oller found that the more complex the text in syntax and lexis, the better the device discriminated between good and weak students. The procedure Oller suggests is to read the passage through once, to re-read it with pauses at phrase or clause boundaries at normal speed, and finally to read it once again to allow students to make corrections.

The use of composition writing as a means of assessing proficiency has also been advocated, but the problems involved in marking such an item seem to out-weigh the advantages of the test. It has been found that impression marking by two or three markers has ensured a degree of objectivity, but the inefficiency involved hardly seems to be justified, particularly as the device does not reveal any information about the student's proficiency that cannot be obtained from a cloze test or a dictation.

4.2.1 The advantages of integrative skills tests.

It should be clear from the train of argument in this paper that the chief advantage of the integrative skills tests is that it appears to test a student's overall proficiency in communicative situations. As Oller points out, those who favour communicative testing reject the notion that "If you can get across 50,000 (or some other magic number) structural items, you will have taught the language." (p.198) It follows from this that however many discrete items are tested it is not possible to get an accurate view of a student's ability to communicate. The brief examination of the strategies involved in the completion of a cloze test indicated that the device tests the student's sensitivity not only to syntax, but to numerous other factors such as an appreciation of text forming cohesive ties, appropriacy and the communicative function of language.

Apart from the problems involved in the training of assessors for the oral interview technique described above, the construction of integrative skills tests is both easy and economical, and they have all the advantages of ease of marking that are associated with discrete item tests. The only disadvantage of integrative skills tests is that they do not isolate specific points in which a student is deficient. However, as was argued above, the identification of such areas of difficulty is best carried out by the teacher in the classroom.

5.0 The proposed test.

In this final section I will suggest the test devices which will be incorporated in the entry test for the course described in section 2.0., keeping in mind Heaton's comment that a proficiency test assesses a student's ability "with reference to a particular task which he will be required to perform".

The test will comprise a reading cloze test, an aural cloze test, an oral interview as designed by Walters, and a dictation. The reading cloze, (see Appendix), is taken from page 77 of The Population Explosion - an Interdisciplinary Approach , a text book used in the Open University Social Science Foundation Course. The text is approximately 450 words long and, after a short introductory passage to familiarize the student with the register of the text, every seventh word is omitted. In scoring, only exact words will be accepted. This is contrary to the recommendations made by Oller, but it would seem that such a method of scoring eliminates problems of agreeing on what is acceptable, and it should differentiate more delicately between good and weak students. The text is fairly complex, but it has been chosen because it reflects the sort of text the students will be required to read in their academic studies and because its content is sufficiently non-technical to be unlikely to favour any particular student. The time allowed for the completion of the test will be 45 minutes.

The aural cloze test, (see Appendix), is a spoken text taken from an Open University programme on the use of university libraries. Again, this particular text was chosen for its content matter, which is of relevance to all students, and for its level of difficulty. The procedure to be adopted will be that suggested by Templeton, i.e. the test will be recorded and conducted in the language laboratory. The students will hear the complete text once, read at normal speed. It will then be re-played with every fifteenth word replaced by an electronic bleep. A break of four seconds duration will occur at the end of convenient sense groups. Some of the 'bleeps' occur at the beginning of a sense group, but the groups are sufficiently short to avoid problems of short term memory. In all, thirty words have been omitted from the text. The test will be completed in the time it takes to play and re-play the tape, approximately 15 minutes.

In pre-testing it was found that this particular device was fairly taxing for native speakers. It is suggested, therefore, that any acceptable word should be scored. The oral interview will be conducted in the manner described above. Each interview will last for five minutes. To ensure greater objectivity in scoring, the assessor will stop making the record of the interview as soon as five minutes has elapsed, even if the conversation continues for a short time.

The dictation, (see appendix), is taken from The Waste Makers by Vance Packard. Again, the subject is of general interest and the

complexity of the text should be sufficient to discriminate between the weak and the good students. The procedure will be that described above in relation to dictation tests. The student scripts will be marked out of 20 with 1 point deducted for any mistake. In the cloze tests one mark will be awarded for every correct or, in the case of the aural test, acceptable entry. The interviews will be graded according to the record plotted onto the matrix.

Spolsky remarks that if integrative skills tests are successful in assessing a student's underlying competence it should be unnecessary to test receptive and productive skills separately. It is argued that the testing devices described force the student to adopt a strategy of analysis-by-synthesis which involves at once both productive and receptive skills. Higgens (1976), on the other hand, found that, when tested in all the different skills, students achieved better scores on some sub-tests than on others, indicating that their reading ability was superior to their speaking ability, or vice-versa. Thus, the theory does not match at least some of the empirical evidence. Perhaps the question posed will be answered when the relevant data has been obtained by administering the test.

References:

Chomsky, N. (1957) Syntactic Structures. Mouton. Le Hague.

(1965) Aspects of the Theory of Syntax. Cambridge Mass. MIT Press.

Chomsky, N. & Hall, M. (1968) The Sound Pattern of English. New York. Harper & Row.

Fries, C.C. (1945) Teaching and Learning English as a Second Language. Michigan: University Press.

Harris, D. (1969) Testing English as a Second Language. New York. McGraw-Hill.

Heaton, J.B. (1975) Writing English Language Tests. Longman.

Lado, R. (1961) Language Testing. Longman.

Oller & Richard (1973) Focus on the Learner. Newbury House. Rowley, Mass. 1973.

Pride & Holme (1972) Readings in Sociolinguistics. London. Penguin Books.

Sinclair-Rogers (1977) The Communicative Needs of some Overseas Postgraduate Students. English for Academic Purposes. Cowie & Heaton (eds.) BAAL/SELMOUS.

Spolsky, B.(ed.) (1969) Language Learning. Special issue on Testing.

Spolsky, B. (1976) Proceedings of the 4th International Congress of Applied Linguistics, Vol.3.

Skinner, B.F. (1957) Verbal Behaviour. New York. Appleton-Century-Crofts. 1957.

Taylor, W.L. (1953) Cloze Procedure: A New Task for Measuring Readability. Journalism Quarterly, 33. pp.42-48.

APPENDIX

Reading Cloze

Instructions: Read this passage and write the words you think are missing next to the correct number on your answer sheet. Write only ONE word for each gap.

Economic Growth and Population Growth.

Malthus' first proposition was that as the standard of living increased population would also increase. Population growth of course is a function of both the birth rate and the death rate. A high birth rate combined with an equally high death rate produces a similar growth in population (either positive, zero or negative) as a low birth rate combined with an equally low death rate.

The Birth Rate.

Taking the birth rate first, Malthus (1) that when incomes rose, people could (2) to marry earlier. The effect of (3) would be that more of a (4) childbearing years would be spent in (5) and, therefore, she would be exposed (6) the risk of having children, and (7) societies with which Malthus was (8) would, he asserted, have more children, (9) if she married later.

There were (10) of this approach in Malthus' own (11). Nassau Senior (1790-1864) argued that the (12) for a higher standard of living (13) as strong an influence on an (14) behaviour as the desire for procreation. (15) tasted a higher standard of living ((16) in the modest form of more (17)) a man would attempt to protect (18) and improve upon it. Thus, a (19) in living standards provided an automatic (20) check to the growth of population.

(21) a number of theories have been (22) forward arguing that economic growth produces (23) change in attitude to birth control. (24) that we are talking about a (25) in attitude for it is argued (26) ignorance of contraception techniques is not (27) basic problem although its removal helps (28) attitudes have changed. Even primitive societies (29) a knowledge of basic contraception techniques (30) as that described in the Bible. (31) problem is whether people wish to (32) birth control. Economic growth may cause (33) change of attitude in a number (34) ways. It may involve a movement (35) a rural to an urban environment (36) with it a weakening of religious (37) that may have associated with attitudes (38) birth control. The political and industrial (39) of women has led to (40) awareness of other functions in life (41) child-bearing and rearing.

Professor Blaug (42) that these socio-economic factors with (43) to having children, can be formalised (44) a cost-benefit analysis. A cost-(45) analysis weighs up all the relevant (46) and all the relevant benefits of (47) particular action and compares them, to (48) whether the outcome is a net (49) or a net benefit. The higher (50) net benefit the greater is the chance of the action being undertaken.

Aural Cloze.

Instructions: You will hear the passage twice. On the first occasion the passage will be complete. Do not write anything. On the second occasion every 15th word will be replaced by a 'bleep'. Write down the word you think has been replaced next to the correct number on your answer sheet. If you do not know the word draw a line next to the no.

This programme is about university libraries.//When you go to your summer school you'll have the chance to use the library of the university you're visiting.//Don't worry if it seems larger or more complex than other libraries you've previously used//but remember you cannot afford to spend too much time finding your way around.//_____ need to go into the library//and as quickly as possible find the___ you want//I hope this programme will help you do just that//there is_____ such thing as a typical university library//they are as varied as the universities_____// the libraries of some of older universities may have over half a million _____//these are comprehensive and scholarly collections built up over many years// the libraries of_____ more recent universities may have smaller but equally wide ranging collections of books//while_____ libraries of the technological universities//are likely to be particularly strong in the pure_____ and social sciences//a university library has to meet the needs of both its_____ and its teaching and research staff//this means that you can expect to find books_____ from undergraduate texts to advanced research monographs//but usually only in the subjects that_____ taught or studied there//100,000 new books are published in English every year//no_____ library can afford to buy more than a fraction of them //so don't expect_____ find every book you want in the university library you're using//you are likely _____ find the library near the centre of the university campus//some universities have one_____ library serving the whole university//this is the case here at Bath//a university_____ is like a department store//like the public library it is mostly self-service_____ guides and signs to help you find your way round//this library has a_____ plan for each floor//this is the entrance floor//just inside the door is _____ loans desk where books are borrowed and returned//nearby are the enquiry desks where_____ go to ask 'how do I find this book//or where do I find_____ information//on this floor too are books which will be of use to all_____ of the library whatever their individual subject interests// on the quick reference shelves are_____ directories, encyclopedias, language directories, trade directories, biographical dictionaries like WHO'S WHO and atlases//in_____ area too are the bibliographies//these are guides to what books have been published_____ what subjects and by what author//we will return to these later//also on_____ entrance floor and this is typical of university libraries//are the catalogues of the_____ that are in the stock of this library//some library catalogues are now in _____ form of computer print-out//or computer-produced microfilm for which special machines are_____//but most catalogues are on cards filed in banks of drawers probably no different_____ what you're used to in your public library//

Dictation.

Instructions: The passage will be read to you twice. On the first reading LISTEN. The passage will then be read again with pauses. Write down what you hear.

Today,//the average citizen of the United States//is consuming twice as much//in the way of goods//as the average citizen consumed// in the years just before the Second World War.//Nearly two-fifths of the things he owns//are things that are not essential//to his physical well-being. //They are optional or luxury items//and there are signs//that physical possessions//are becoming too plentiful//to accommodate comfortably.

Testing Communicative Competence in English as a Foreign Language with Students of Natural and Social Sciences.

S. McDonough

University of Essex Department of Language and Linguistics
Occasional Paper No.21 pp.83-96

In this paper I shall outline some reasons for the development of a new type of oral production test, describe a test procedure using Upshur's (1971) method with graph material like that found in any specific or social science text, and analyse some of the problems that this procedure poses.

The need of a test of communicative proficiency

There is an accelerating trend at the moment towards the use of language teaching materials that involve communication problems. Within the language lesson, students are encouraged to attempt to create utterances that are novel to them, and give their peers and the teacher information that is genuinely new, on the grounds that language is essentially creative and that using it to express something new to the hearer is at once natural and stimulating to the learning process. Within the language course, there is increasing use of syllabuses that group and sequence the new language features to be taught according to functional and/or notional classifications rather than structural ones (Wilkins, 1976). This change in language teaching strategies has been motivated by a variety of considerations, such as a deeper understanding of sociolinguistic variables, and the existence of consumers of language courses who have quite well-defined (but not necessarily restricted) aims in acquiring the new language. The new trend has brought about a change in the terms in which students' attainment is expressed. Attainment can now be stated in terms of, for example, the student's ability to review, summarise, explain, present a conclusion, greet, transmit information, etc., in language X, and those abilities can be related fairly directly to the specification of the students' intended use of the language. Previously, attainment was in terms of an ability for example, to form and use structures 1 - n in language X. The change in aims and the concommittant change in the definition of attainment means that assessment procedures have to be revised and, if necessary, new ones adopted; tests of attainment or proficiency have a well-known backwash effect and if they are inappropriate, the changes in method, however well they may be intrinsically justified, will be slowed down or reversed. The older type of test may not be sensitive to the new skills being included in the curriculum. Thus, students' mastery of these skills may not be reflected in the assessment based on the conventional tests; and the newer types of course may not receive adequate formative evaluation, i.e., feedback on the success of the course in fulfilling its own aims.

Conventional oral production tests tend to be either highly artificial and easily scorable, or realistic and difficult to score. Structure tasks such as reading aloud or oral drills can have a place in a battery of tests but are essentially tests of performance in the sense that they measure performance of some previously learnt routine or exercise type (e.g. structure manipulation or the production of a reasonable-sounding intonation). They do not test spontaneous communicative use of the language. They are relatively easy to make objectively scorable, although the criteria for acceptable pronunciation in reading aloud may be difficult to state. Less objective, but still

controlled, tasks, are the various kinds of situational response exercises, where a short text describing a situation is read to the candidate and an appropriate response demanded (e.g., "You want to order a book which is not in stock. What will you say?); and descriptions and story narrations based on pictures displayed to the candidate. Some examples of these will be found in Templer and Nettle (1975). Situational responses may be difficult to score; in particular, an original response which deviates from the examiner's list of acceptable responses may be penalised, although it might achieve the desired aim in real life. Picture descriptions can reduce to vocabulary tests unless the pictures are well chosen and sufficiently ambiguous for a good sample of the candidate's language proficiency to be provoked. The most naturalistic, but also the most difficult type of oral test is the oral interview. Of this, Heaton, (1975) says:

"Like many other examinations of oral production, the scoring of the oral interview is highly subjective and thus sometimes has only low reliability. In addition, the performance of a student in a particular interview may not accurately reflect his true ability." (P.94)

The sources of unreliability are the candidates' interests, and degree of nervousness, the ill-defined categories of the scale used for rating the candidates' performance, and very often the boredom of the interviewer. It is naturally difficult to ensure the same degree of difficulty of communication problem if the students' own interests are involved. A candidate may score highly if a subject of interest to him is broached early in the interview, another may score unfairly low if he is unable to get on to a subject of interest to him until late. At the same time, candidates' own interests are likely to determine effort in preparing for the interview; thus the score may be unrepresentative of his general level of competence anyway. The measure used is typically a rating scale such as that proposed by Harris (1964); while it is possible to train observers to achieve a high degree of agreement in the assignment of candidates to categories, it is rarely possible to analyse the category descriptions and relate them to other tests for cross-validation. Thus, the actual criteria used by even trained observers are not available. It is likely that a variety of criteria are used, perhaps at different points in the scale. Thus while ostensibly rating only the language, an observer may be rating success at solving the communication problem as well, or even the intelligence of the candidate.

Many of these difficulties have been recognised for a long time, and this, as well as cost, may account for the relative unpopularity of oral production tests as against pencil and paper tests. Any oral production test must conform to the normal criteria of assessment instruments, namely validity, reliability, cost, feasibility, and discrimination.

The 'Productive Communication Test' (P.C.T.)

From the foregoing criticisms of traditional tests it appears that a good communicative test should have the following characteristics:

(a) the test should pose a communication problem: that is, there should be a realistic information gap between candidate and examiner to be bridged by the candidate.

(b) the content to be communicated should be controlled and therefore held constant for each candidate.

(c) the measure should reflect actual success in solving the communication problem posed, and not merely rate superficial features of the language employed.

(d) the test should be simply scoreable.

It might seem that characteristics (a) and (b) are contradictory, but a technique has been suggested by Upshur (1971) and Palmer (1972) which has all four characteristics. In this method, sets of minimally different pictures were designed so that each picture had a unique combination of attributes of binary features. For example, for a picture set about someone in the street, there might be two features: sex and weather. A set of four pictures is necessary to portray the conjunctive possibilities here:

(1) a man in the rain

(2) a woman in the rain

(3) a man in the sun

(4) a woman in the sun

The candidate's task, faced with these four pictures, is to choose one and describe it unambiguously to the examiner, who has an identical set. Thus the examiner does not know exactly what will be described (characteristic (a)), but knows the alternatives (characteristic (b)). The candidate has to find language to express the two attributes which define his chose picture. The measure used is whether or not the examiner selects as the one described the picture intended by the candidate (characteristics (c) and (d)). In the simplest and most reliable form of the test, no other communication between candidate and examiner (e.g. supplementary questions) is allowed (Palmer, 1972). As normally used, the measure actually involves time as well as accuracy, either the average time taken to transmit pictures from n sets, or the number of successful communications within a certain time limit. Both Upshur and Palmer found that this type of test correlated quite highly with more traditional types of grammatical accuracy tests. This might be expected on the grounds that the more easily accessible the means of expression are to a testee, the quicker he will be able to communicate, whereas the more circumlocutions he has to make, or the more false starts he makes, the longer it will take. However, even a candidate with severe language problems can probably indicate his choice somehow, if given a very long time.

Scientific applications

One of the language tasks that faces students in English-medium higher education frequently in the fields of natural and social science is that of talking about graph material and explaining the meaning of a visual display, in a seminar or tutorial presentation. This is sometimes known as 'channel conversion' - translating from a visual to a verbal mode. Upshur's procedure can be adapted to this situation moderately easily, providing a direct test of one of the students' most necessary communicative skills.

Material can be prepared relatively easily with either two features (4 pictures per set) or three features (8 pictures per set) using the following types of graphic:

(a) straight line graphs

(b) curvilinear graphs

(c) multiple line graphs

(d) pie charts

(e) histograms

(f) ideographs

(g) distribution maps

and probably others, e.g. flow charts.

There are some problems in designing sets of such items. For example, no picture should be more easy to describe than any other within a set. It can happen that one graph has a shape which is more memorable than the others or is more easily glossed by an analogy which enables it to be referred to unambiguously without specifying the distinctive features (e.g. "the one that looks like a hat"). If this is not obvious by inspection, pretesting will reveal such items, or if any item is being chosen disproportionately frequently. Another problem is that of the simulated context. This enables the graph to be interpreted in terms of the real world but probably results in words on the page in the form of a legend, which may introduce comprehension problems. Also, all the pictures must be equally plausible for that context. A graph can be talked about in the abstract, as representing relationship R between N variables, and post-graduates and undergraduates with relevant qualifications in natural and social sciences are used to talking in this way in their own language, but naturally have problems in the foreign language. Some examples of sets of graph materials are presented in the appendix.

A version of this test has been pre-tested with 24 members of a group of Spanish-speaking post-graduates in social sciences, mainly economics, industrial studies, finance, etc., learning English prior to starting postgraduate courses in British Universities and Polytechnics. This group was familiar, by training, with interpreting graph material in Spanish. They were tested at a time when for most of them, their English course had not changed from being an introduction to basic English proficiency to Study Skills and ESP, except for some number work and a transitional programme. The British Council, who were administering the English Teaching programme at national level, gave a fairly traditional English proficiency test at about the same time, and permitted a comparison of results. The P.C.T. was administered alone with the student, recording

the choice, the time taken to communicate, and the actual words (on cassette). Instructions were given in Spanish. The two measures resulting - success and time - were converted into a notional time by assuming that failure incurred a penalty of five minutes, and a simple average was taken over all the sets. Rank order correlation between P.C.T. and the British Council Test overall and between P.C.T. and the four components of the British Council Test are given in Table 1.

Table 1

P.C.T. and British Council Test Overall Rank	$r_s = 0.4^*$
P.C.T. and Translation subtest	$r_s = 0.36^*$
P.C.T. and Cloze subtest	$r_s = 0.35^*$
P.C.T. and Structures and Vocabulary subtest	$r_s = 0.46^*$
P.C.T. and Explanations subtest	$r_s = 0.25$

$^* = p < 0.05$

There is, from these figures, a reassuring degree of agreement in the rank orders of proficiency between the P.C.T. and the more traditional test, as Upshur had also found. It does not correlate with the components as highly as they correlate with each other, the inter-correlation coefficients varying between 0.4 and 0.6. This is not too surprising, since it is a rather different type of activity. The highest correlation here is with the Structure and Vocabulary (multiple choice) component: this confirms Upshur's finding, and perhaps indicates merely how much easier grammatical English is to understand than ungrammatical English.

Both the Cloze and the Translation are productive tests in varying degrees, and this probably explains the significant correlation here.

The only non-significant correlation, the low one with the Explanation task, is on the face of it disappointing. In this written task, short three- or four-sentence explanations in English were sought for simple problems. However, the choice of problems may have been inappropriate for this particular group - they were mainly simple physical problems (balance, heat, etc.) - and the marking scheme may have contributed some unreliability. It is difficult to establish agreed criteria for "powers of argument and exposition" (quoted from the rubric of the test) and divorce them from language problems. It may be, therefore, that the low correlation with this component is due to low reliability of the component.

The above results give some clues as to the validity of the P.C.T. in this register. Unfortunately it was not possible to take any other measures. It is possible, for example, that scores on the P.C.T. would vary with some factor like intelligence as well as language proficiency measures. No measure of this was available. On the other hand, the testees were all graduates selected for postgraduate training, and one would not expect very wide differences in measurable intelligence among such a population.

Problems

Firstly, the aim of a test such as this is to measure a particular communication skill in a foreign language. The sense of 'communication' here is obviously an information-theoretic one: a certain number of binary choices of data must be transmitted somehow in English. It is equally obvious that a 'good communicator' is often taken to be someone who can do rather more than this; perhaps mix easily, or persuade, or use appropriate social formulae. The P.C.T. in this form does not test these wider aspects of communication, and no test so far proposed can, either.

Secondly, time is involved in the measure. Although this is true of any test, even those in which the time allotted is supposed to be enough for even the weakest candidate, in the P.C.T., the time taken to communicate is a crucial component. Unfortunately, time may vary with factors that have nothing to do with language proficiency, like the candidate's normal working speed, his nervousness, or his understanding of the task.

Thirdly, the total time for the test may be in the region of 20 minutes depending on the number of sets used. This could prove expensive. Tape-recording the responses for later evaluation can allow simultaneous administration to a group.

Fourthly, candidates develop certain communication strategies which may affect the time, e.g. repetition and rephrasing. These may be seen as requests for assurance, information, or guidance. In fact the tester, in this method, does not respond to these or engage in any conversation once the picture set has been revealed, but the candidates do not always remember this. The transcripts in the appendix exhibit several of these communication strategies. Non-verbal interaction is more difficult to exclude. The best seating position is one in which there is no face-to-face communication: for example, both facing the same way.

Lastly, communication success is not only a function of the candidate's performance, since the measure taken is the time until the examiner has understood, that is, chosen the alternative intended by the candidate. The contribution of the examiner to this measure is divisible into two factors. The first factor is his knowledge of the material: for a trained examiner this is constant for all candidates. The second factor is the degree to which he/she will compensate for the candidate's poor English. This might be a case of recognising common errors (e.g. Spanish "j axis" for "y axis") and translating, or a more serious issue of matching his allowances to the degree of compensation habitually given by the candidate's eventual real audience, which might be his university teachers. This second factor may, of course, work to reduce the time to success. This question of compensation, although important for all forms of oral test, has not yet been widely investigated systematically.

Conclusion

Despite the real problems with any comparatively untried type of test, I suggest that it is worth persevering to find the validity and utility of productive communication testing, since there exists here the attractive possibility of an oral test which is

(a) standardised

(b) controlled

(c) objectively scorable

(d) flexible

(e) relevant to students' experience

(f) relevant to students' future language needs.

Acknowledgement

I should like to acknowledge the help I have received from the British Council, and from staff and students of the King's School of English, Bournemouth; the Centre for Applied Language Study, University of Reading; the EFL Unit, Dept. of Language and Linguistics, University of Essex; and Hilderstone, Broadstairs.

References

Harris, D.P. 1969, Testing English as a Second Language. McGraw-Hill

Palmer, A. 1972, Testing Communication. International Review of Applied Linguistics, Vol.X, No.1.

Templer, J.C. & Nettle, K. 1975, Oral English Proficiency Tests Heinemann

Upshur, J. 1971, Productive Communication Testing, in Perren, G.E. & Trim,J.L.M. (Eds.) Applications of Linguistics, Cambridge University Press.

Wilkins, D.A. 1976, Notional Syllabuses. Oxford University Press.

Heaton, J.B. 1975, Writing English Language Tests. Longmans.

Appendix

This appendix contains two examples of the pre-test materials, together with the transcripts produced by six candidates, and an English translation of the instructions, which were given in the candidates' L_1, Spanish. Students A,B,C were overall high scorers, students D,E, & F were in the middle range, and student G was a low scorer, in terms of their average time compared to the whole group. On the basis of the candidates' responses, a number of revised item sets have now been constructed.

Instructions:

This test is designed to measure your ability to communicate in English. You will be shown several sets of pictures and graphs; the members of each set are quite similar to each other, and are presented simultaneously. Your task consists of the following:

(1) Choose one picture, but do not say which you have chosen, nor name the letter on the bottom of it.

(2) Make a note of the letter on the picture or graph.

(3) Describe the picture which you have chosen; do so in English, and in such a way that it is clear which picture you are describing.

(4) Try to do so as quickly as possible.

The first set of pictures is a practice set. Remember the most important points are:

(A) Choose, but do not name the letter

(B) Note down the letter

(C) Describe the picture.

Transcripts:

(The letter in brackets after each transcript indicates the picture or graph the candidate was attempting to describe.)

(1) The simple line graph.

Student A

This very easy. OK. you know that we have a line vertical and one horizontal and this one is er the more horizontal than the others one. And in the line y the line begin on the point 20 and it is more horizontal than the others points the others graphics. I'm sure that you know what is. (D)

Student B

About 20 and up about 45 or 40 right up. (A)

Student D

I have a statistic figures. The is statistic figure is a fartesian system. There is a line that shows some numbers and another line x line that shows the same number. In this Cartesian system is represented a line who shows a tendency increasing. In bothe case the tendency is improving. For example in the y line in the x line the tendency begin in 40 and achieve the most quantity until 60 and in other line in the x line the tendency begin zero and achieve a grade of 100 In this case the improving of the tendency is majority is major in the x line and the x line the tendency is increasing but this increase little amount is little amount if I have to compare in the increase in the x line (B7)

Student E

There is a line of one function the line begin where the funct[illegible]n have 2[illegible] in (j) and then have a little climb I think. This line is no happen (D)

Student F

In this picture I can see one diagram of the Cartesian coordinates of y the line starts the 20 the line is straight no is curved is straight. It is straight at 20 and is oblique The grade is 0 and 10 in every case (7)

Student G

The graphic have two lines, one line horizontal and another line vertical. On the left I can see many number. On the right I can see many number for example 40 20 30 60 On the... I can see the letter j is all.

(2)

Pie Chart (car imports)

Student A

OK In this graphic we can see this country import more material and more cars from Germany and also from France and Japan but the proportion of Japan is a little more than France and I think that this graphic or this country is the country that import more than the other graphic than the others country and if you want I can tell you more the Germany and the imports from Germany are situated in the superior part of the circle (B)

Student B

I think in this US don't import cars as France Japan and Germany but the production cars is similar in Germany. The production car in United State is same import cars Germany and France I think import more cars than Japan in this graph Oh yes US don't import cars - produce all his cars (C)

Student C

Well in the diagram I see that the car imports from US is nothing and the car important from France is in second place. No. Excuse me I will say the car import from Japan is in the first place because it is the most big area in the circumference in the second place I will say Germany and the third place France. I will conclude that the people import more car from Japan than the other country.

(D)

Student D

Here is a statistic figure represented by a circumference The figure represents the quantities of car imports from some countries eg it shows that Britain import from France certain quantities of cars but this quantities is less than the quantities import from Germany and Japan and the United States. The quantity of cars import by for Britain from France is approximately the 30% and from Japan is approximately the 45% This quantities of course majors that the quantities import from Britain, the United States and France.

(D)

Student E

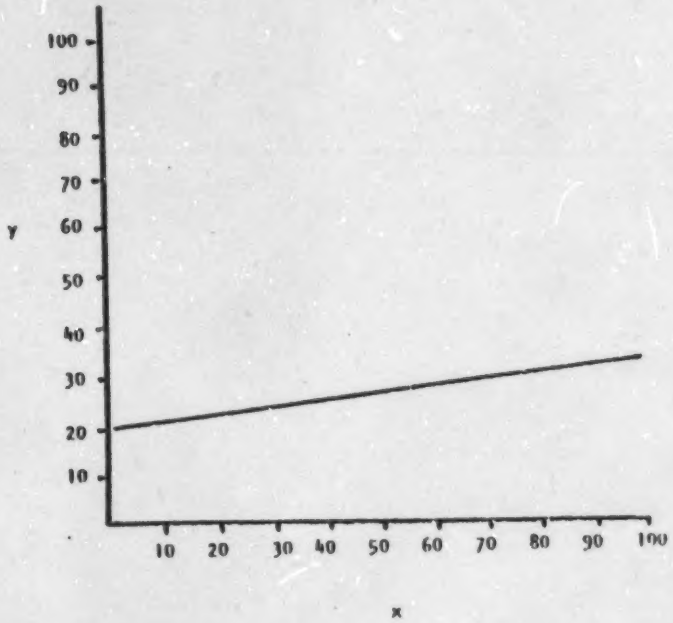
The graph is a representation of the importation of cars from countries - France Japan Germany United States is the country with less car importation and G German is the country with most importation I think in % German is the first, the US the less, Japan is the third, France is the fourth in % of importation of cars.

(?)

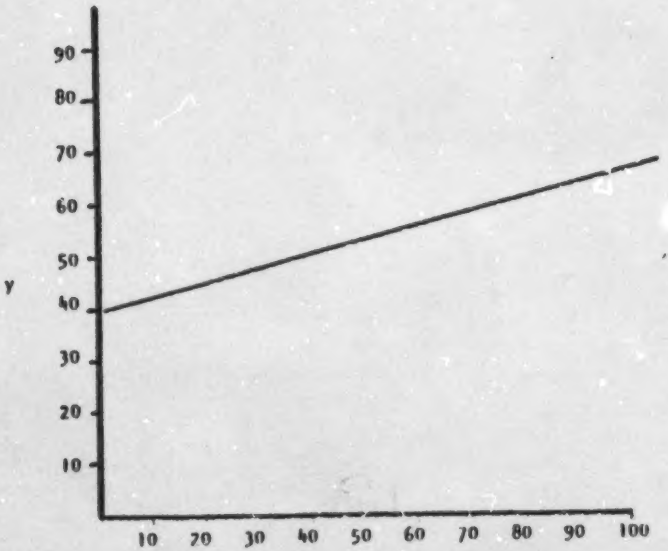
Student F

Car imports. I can see in this picture one diagram in the form of cakes. It shows the car imports. I can see the principal country import is Japan because the the more part the cake is this country second is German after is France and the more little is America.

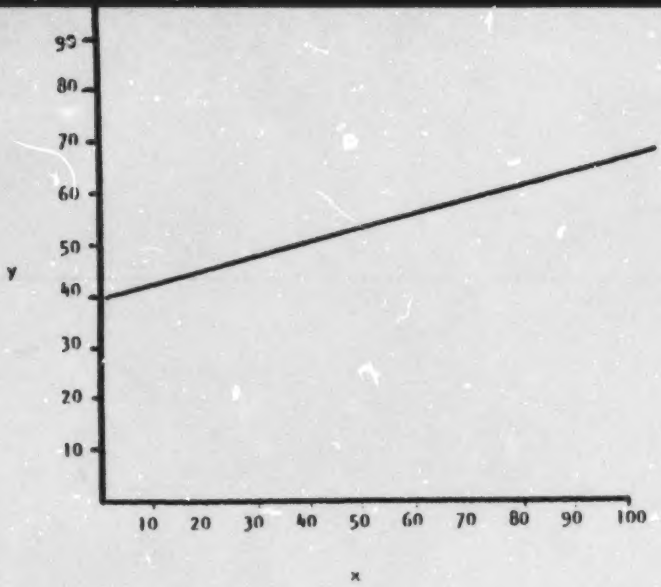
(C)


100
90
80
70
y 60
50
40
30
20
10
10 20 30 40 50 60 70 80 90 100
x

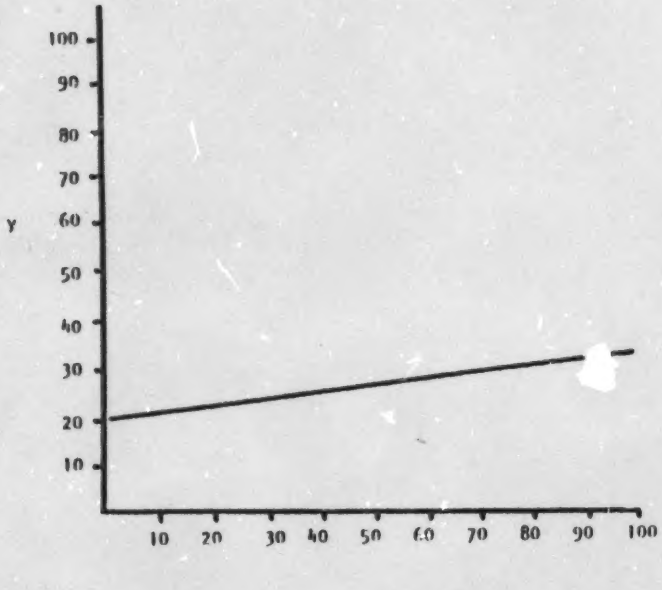
A


90
80
70
60
y 50
40
30
20
10
10 20 30 40 50 60 70 80 90 100

B

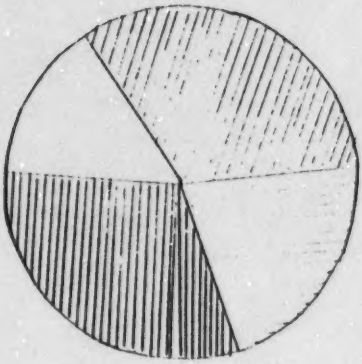

90
80
70
60
50
40
30
20
10
y
10 20 30 40 50 60 70 80 90 100
x

C


100
90
80
70
60
50
40
30
20
10
y
10 20 30 40 50 60 70 80 90 100
x

D

Car Imports



From

France

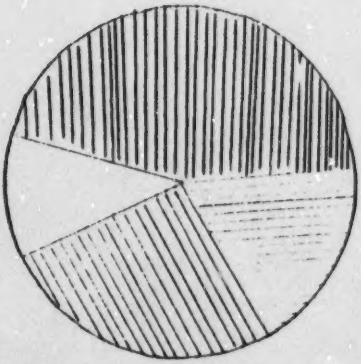
Germany

Japan

USA

A

Car Imports



From

France

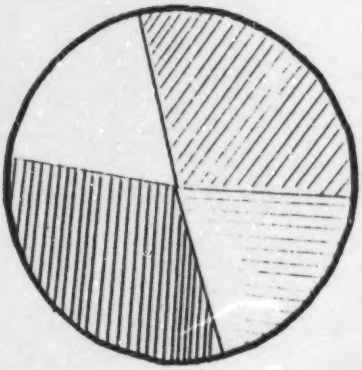
Germany

Japan

USA

B

Car Imports



From

France

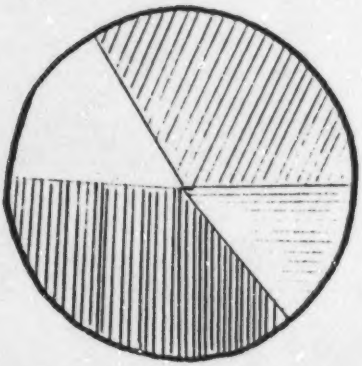
Germany

Japan

USA

C

Car Imports



From

France

Germany

Japan

USA

D

A proposal: Use of analysis by synthesis to investigate naturalness of second-language pronunciation.

K. Robinson
D. Rekart
R. Daniloff

University of Essex Department of Language and Linguistics
Occasional Paper No.21 pp. 98-113

0. Despite rapid advances in experimental phonetics and speech perception in recent years, there has not been a corresponding advance in the teaching and testing of foreign language pronunciation. In this paper we intend to sketch an avenue of research, utilizing knowledge gained from experimental phonetics which might provide for a theoretically satisfying way of testing foreign language pronunciation.

1. There are two central issues to be confronted if we hope to investigate empirically proposals for the teaching and testing of pronunciation: (1) how may we determine what to teach, and (2) how may we know when we have succeeded in teaching it.

1.1 Many proposals have been advanced as to what aspects of pronunciation should be taught.

(1) According to those advancing an articulatory approach (Reichmann, 1967), it is articulatory movements of the "visible articulators.... used with exaggeration" which are fundamental to the teaching of pronunciation.

(2) Adherents to the auditory approach (Celani 1967, Lebel 1969, DeGreve 1968) view preliminary instruction in auditory perception as essential because correct production implies correct perception.

(3) On the other hand, the motor theory approach to teaching pronunciation (Trocme 1974) is based on articulation since the acoustic signal and its interpretation is an act of reference back to articulation.

(4) Central to the verbo-tonal method of teaching pronunciation is the degree of "tension" or articulatory energy that sounds have which allows them to be ordered hierarchically according to degree of difficulty (Renard 1971). In addition, suprasegmental factors such as intonation, tempo and intensity should be included with the items to be taught in a pronunciation course.

(5) The approach to teaching pronunciation which has actually attempted to answer most clearly the question of what is to be taught is known as contrastive analysis. The fundamental assumption underlying contrastive analysis is that individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language to the foreign language - both productively when attempting to speak the language (Fries 1945; Lado 1957). Therefore, what should be taught are those points where the systems of the target and native languages differ.

The hypothesis advanced by proponents of contrastive analysis is supported experimentally by R. Scholes (1967, 1968), who found that

the English categorization of synthetic vowels by non-native speakers paralleled their native-language categorization (in so far as the native-language categories had phonetically similar English counterparts).

Although, as we have seen above, a variety of answers have been given to the question of what to teach, little attention has been addressed to the issue of choosing among competing proposals on some nonarbitrary basis. Yet this is the heart of the entire matter and it is reasonably clear how this issue can be approached. First, it is obvious that the judgement of native speakers of the language as to the quality and/or intelligibility of one's pronunciation must be the ultimate authority in evaluating it. Should the native speaker find no difference between one's pronunciation and his own, then clearly one has little further room for improvement. But if the native speaker has difficulty understanding what has been said, or if his attention is constantly distracted from the message to the poverty of its realization, then this pronunciation must be assigned a lower value.

In an important recent article, Dimitrijević and Djordjević (1971) have addressed the question of the subjective evaluation of the pronunciation of English by Serbo-Croatians. They used four categories of examiners: secondary school teachers, phoneticians, native speakers of English and senior university students of English, and found the following results. The native speakers were the most uniform in their subjective assessment of the pronunciation, followed by the phoneticians, English teachers and students, respectively.

If untrained native speakers are the most uniform group of examiners (as well as being the obvious target audience for one's pronunciation), then it is crucial to investigate what aspects of pronunciation affect their evaluation. It is difficult to see how there can be any other basis for determining what aspects of pronunciation should be taught.

2. We are dealing, then, with a program of research in the perception of the native speaker, and only secondarily with the production of the student. Again, the basic outline of such a study is reasonably self-evident. Let us first sketch an idealized program and then ask whether it is possible to carry it out in the current state of the art.

Assume first the existence of two stimuli which are (1) the pronunciation of a sentence by a native speaker and (2) the mispronunciation of the same sentence by a student of the language. The two stimuli are presented pairwise to listening crews (native speakers) who are asked to decide (a) which has the better pronunciation and (b) which is easier to understand.

Assume next that it is possible systematically to vary the phonetic properties of the two stimuli in a controlled fashion so that one can leave stimulus (1) unchanged except for modification to exhibit some single aspect of the pronunciation of stimulus (2), e.g., all the voiceless stops have been substituted, or the rhythm characteristics have been substituted, etc. One then has both the original stimuli and one or more derived stimuli. This entire set of stimuli is then presented pairwise to listening crews for the same kind of ranking judgements given above.

If one can systematically carry out the process of deriving stimuli over the entire gamut of phonetic characteristics of the two languages, then the result of the ranking tests would provide a clear picture of which phonetic characteristics and combinations thereof cause one to be perceived as having bad pronunciation. These are precisely the features which must be taught.

The difficulty lies in the systematic, controlled alteration of the stimuli. It is simply not possible for a human speaker, however gifted, to articulate sentences with such close control that he reproduces exactly his pronunciation of a former utterance with the change of only a single parameter. A solution to this problem would be as follows. The two original stimuli are produced as before, but instead of using them directly the same two sentences are used as models for speech synthesis using a digital computer with a synthesizer program (e.g. Klatt 1977). Once the two original stimuli are approximated as closely as possible with two synthetic stimuli, one then has not only the two stimuli but also the sets of parameters which were used as input for the synthesis. It is then possible systematically to recombine the two sets of parameters in a very large number of ways, each producing a completely controlled intermediate stimulus for use in the ranking tests.

3. Theoretical Rationale

Introduction

Digital synthesis of speech is a well-established research technique in speech perception (see Flanagan and Rabiner 1973 for a comprehensive review, or Fant 1975), and much of our knowledge of how speech is perceived is derived from studies of listeners' reactions to computer-synthesized speech signals (Studdert-Kennedy 1976). By far, the majority of investigations of speech perception involving synthesized speech have focused on the cues underlying phonemic labelling, perception of suprasegmental cues, and studies of dichotic, two-eared competing signal processing. If one is interested in synthesizing speech utterances larger than the syllable, let us say a phrase or sentence, greatly increased difficulty is encountered since one must take into account stress, tempo, rhythm, etc. etc. In addition to all of the problems associated with segmental and syllabic synthesis. A number of schema for the synthesis of long stretches of speech have been offered (Rabiner 1969; Holmes, Mattingly and Shearme 1964; Fant 1975), but they are far from satisfactory. Nevertheless, it will be the thesis

of this paper that speech synthesis of relatively long utterances is a theoretically promising technique for investigating the "adequacy of pronunciation" achieved by speakers learning a second language.

In order to make our thesis explicable, we must first discuss, in broad outline, some of the principles and practical procedures involved in speech synthesis. As an acoustic system, the vocal tract consists of an acoustic tube, of complex-shape, filled with a column of air. The air column can be excited into vibration and resonance by the injection of high-speed air bursts at the glottis, or by turbulent air streams created by constricting the tube at some point above the glottis by bring the articulators into approximation, or by closing the vocal tract at some point, and rapidly opening the closure, releasing the pent-up air pressure as an acoustic shock wave. In his pioneering work on the acoustic theory of speech production, Fant (1960) convincingly demonstrated that if the investigator knew the cross-sectional shape of the vocal tract at every point along its length, he could predict the transfer function, H(s), of the vocal tract acoustic tube. The "source x filter" model of vocal tract operation suggests that the speech sound P(s) can be viewed as the result of the "filtering" action of the vocal tract upon the input (usually glottal) source of acoustic vibration, i.e.:

$$P(s) = H(s) \cdot S(s) \cdot R(s)$$

where P(s) = complex frequency spectrum of speech sound

H(s) = complex frequency representation of vocal tract transfer function

S(s) = complex frequency spectrum of input sound

R(s) = radiation function of lip/head system

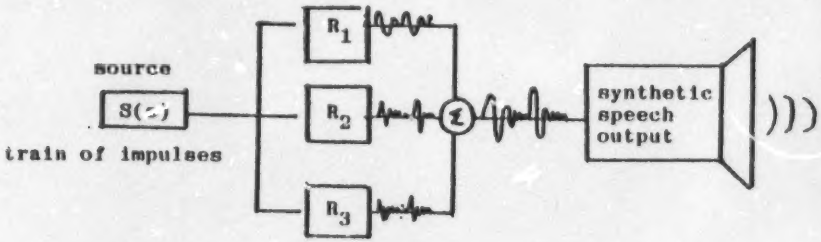
For the vocal tract viewed as a cascade series of acoustic tubes, Fant (1960) demonstrated that the H(s) function could be predicted in its entirety if all pole (resonance) and zero (anti-resonance) frequencies as well as their damping constants were known. Furthermore if the source spectrum S(s) was emitted at the vocal folds and if some assumptions were made about its spectral shape, within limits the frequency spectrum of the resulting speech sound could be predicted quite accurately. Thus, in theory, if all pole and zero frequencies and their damping constants were known, and if the fundamental frequency and spectral shape of the source spectrum S(s) were known, the speech sound spectrum is predictable.

With some effort, the mathematical equations above can be programmed as a set of equations in a digital computer. If the experimenter inserts the appropriate pole, zero, band-width, f_0, and spectrum shape values in the equations, the computer solves the

equations and generates a digital version of the spectrum of the speech sound, P(s). The spectral representation can be converted into a time-varying wave form, P(t), of the speech sound, and by use of a digital/analog converter, the synthesized speech sound can be visually examined, or heard over a loudspeaker. In order to generate a stream of differing speech sounds, the operator of the digital synthesizer must have a table of control parameters which vary over time in a systematic pattern which matches the changes in the speech sound pattern over time. It would be ideal if continuous speech could be synthesized by first synthesizing each separate sound, and conjoining them in some simple way. Unfortunately, in real speech, the coarticulatory, mechano-inertial, rhythmic and rate-dependent interactions between neighboring sounds is extraordinarily complicated.

The vocal tract's acoustic behaviour can be modelled mathematically as a set of equations, or networks of electrical resonators can be fabricated which oscillate in a way analogous to the fashion in which the vocal tract air column oscillates during speech. Such "analog" synthesizers are of two types, depending on how the resonant circuits are connected.

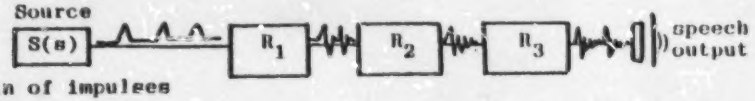
In the parallel terminal-analog synthesizer, the set of resonators are connected in parallel, each being excited electrically from a voltage source of repetitive pulses, analogous to the pulses of the glottal source. The resonator outputs are then added to produce an electrical oscillation which has, when played over a loudspeaker, an acoustic spectrum similar to that of real speech. Figure 1 below shows such a parallel synthesizer, adapted from Flanagan (1957).



Says Flanagan (ibid.),

> 'Each resonator contributes a maximum to the spectrum of the synthesizer output corresponding to a so-called formant of the sound to be generated. For good vowel production, at least three such formant resonances (F_1, F_2, F_3) are usually required ...."

In the cascade terminal analog synthesizer, the resonators are connected in series, the electrical output of each resonator serving as the input to the next one, and so on, as is shown in Figure 2, below.



Train of impulses

As Flanagan has noted, the cascaded (series connection) terminal analog is more realistic in terms of actual vocal tract operation since a series connection of resonators is analogous to the series-connected "tubes" of the vocal tract cavities. In the cascade synthesizer, the "relative amplitudes of the formants in the transfer function, H(s), are determined automatically once the investigator has selected formant frequencies and formant bandwidths. In the parallel synthesizer, the formant amplitudes must also be selected by the experimenter if a naturalistic synthetic sound is to be produced. Thus, the cascade synthesizer is a "more accurate physical analog of the vocal transfer function for vowel production than ..." is the parallel connection. However, for production of consonant sounds, zeroes (antiresonances) must be introduced into H(s), a task which is impossible for the series-cascade set of resonators, but an easy task for the parallel resonator array. "Thus, the cascade synthesizer controlled only by formant frequency information is congenitally incapable of produced consonants such as voiced stops or nasals. There is some evidence that a synthesizer... involving a combination of cascade and parallel resonators might be used to achieve an efficient terminal analog.. " system.

And indeed, the most recently devised digital speech synthesizer, Klatt (1977) offers a combination of series/parallel resonance circuits, provides for antiresonances to be inserted into H(s) for consonant sounds, and is probably as flexible a synthesizer as has been derived to date. To operate such a synthesizer successfully, the operator must specify an extensive list of control parameters which dictate the acoustic properties of the synthesized sound. Shown below, in Table I, is a complete list of the control parameters to be specified for successful operation of the Klatt (1977) digital terminal analog synthesizer. You will notice that certain of the parameters are fixed in value, throughout a long and complicated synthetic speech utterance, whereas others can or must change with each change of phoneme, or must change even within a phonemic-segment. The Klatt synthesizer is designed so that a change in a control parameter can occur every 5 msec., but not in less time. But, since most linguistically and perceptually important speech parameters rarely change that fast, this is not a serious limitation.

Table 1. List of control parameters for the Klatt cascade/parallel formant synthesizer. The second column indicates whether the parameter is normally constant (C) or variable (V) during the synthesis of English sentences. Also listed are the permitted range of values for each parameter.

N	V/C	SYM	NAME	MIN	MAX	SEG Type
1	V	AV	AMPLITUDE OF VOICING (DB)	0	80	C,V
2	V	AF	AMPLITUDE OF FRICATION (DB)	0	80	C
3	V	AH	AMPLITUDE OF ASPIRATION (DB)	0	80	C
4	V	AVS	AMPLITUDE OF SINUSOIDAL VOICING (DB)	0	80	C
5	V	FO	FUNDAMENTAL FREQUENCY OF VOICING (HZ)	0	1000	C,V
6	V	F1	FIRST FORMANT FREQUENCY (HZ)	200	900	V,C
7	V	F2	SECOND FORMANT FREQUENCY (HZ)	500	2500	V,C
8	V	F3	THIRD FORMANT FREQUENCY (HZ)	1300	3500	V,C
9	V	F4	FOURTH FORMANT FREQUENCY (HZ)	2500	4500	V,C
10	V	FNZ	NASAL ZERO FREQUENCY (HZ)	250	700	C
11	V	AN	NASAL FORMANT AMPLITUDE (DB)	0	80	C
12	V	A1	FIRST FORMANT AMPLITUDE (DB)	0	80	V,C
13	V	A2	SECOND FORMANT AMPLITUDE (DB)	0	80	V,C
14	V	A3	THIRD FORMANT AMPLITUDE (DB)	0	80	V,C
15	V	A4	FOURTH FORMANT AMPLITUDE (DB)	0	80	V,C
16	V	A5	FIFTH FORMANT AMPLITUDE (DB)	0	80	V,C
17	V	A6	SIXTH FORMANT AMPLITUDE (DB)	0	80	V,C
18	V	AB	BYPASS PATH AMPLITUDE (DB)	0	80	
19	V	B1	FIRST FORMANT BANDWIDTH (HZ)	40	200	V,C
20	V	B2	SECOND FORMANT BANDWIDTH (HZ)	50	250	V,C
21	V	B3	THIRD FORMANT BANDWIDTH (HZ)	60	300	V,C
22	C	SW	CASCADE/PARALLEL SWITCH	0(CASC)	1(PAR)	C,V
23	C	FGP	GLOTTAL RESONATOR 1 FREQUENCY (HZ)	0	600	C,V
24	C	BGP	GLOTTAL RESONATOR 1 BANDWIDTH (HZ)	100	2000	C,V
25	C	FGZ	GLOTTAL ZERO FREQUENCY (HZ)	0	5000	C,V
26	C	BGZ	GLOTTAL ZERO BANDWIDTH (HZ)	100	10000	C,V
27	C	B4	FOURTH FORMANT BANDWIDTH (HZ)	100	500	VC
28	C	F5	FIFTH FORMANT FREQUENCY (HZ)	3500	4900	VC
29	C	B5	FIFTH FORMANT BANDWIDTH (HZ)	150	700	VC
30	C	F6	SIXTH FORMANT FREQUENCY (HZ)	4000	4999	VC
31	C	B6	SIXTH FORMANT BANDWIDTH (HZ)	200	2000	VC
32	C	FNP	NASAL POLE FREQUENCY (HZ)	200	500	C
33	C	BNP	NASAL POLE BANDWIDTH (HZ)	50	500	C
34	C	BNZ	NASAL ZERO BANDWIDTH (HZ)	50	500	C
35	C	FRA	LOCATION OF RADIATION ZERO	0	100	C,V
36	C	SR	SAMPLING RATE	5000	20000	-
37	C	NWS	NUMBER OF WAVEFORM SAMPLES PER CHUNK	1	200	-
38	C	GO	OVERALL GAIN CONTROL (DB)	0	80	C,V
39	C	NF	NUMBER OF CASCADED FORMANTS	4	6	-

The speech parameters needed to control the synthesizer are generally gathered from measurements made on acoustically analyzed real speech. Thus, from spectrographic data F(f) or oscillographic P(t) data, one can extract formant frequency, formant amplitude, formant bandwidth, f_0, I_0, noise duration, noise-pole frequencies, etc. etc. in abundance, for speakers of various sexes, ages, dialects, languages, etc. The experimental phonetic literature is rich in studies which have determined a wide range of speech parameters across a variety of conditions. Unfortunately, the data are taken or shown only for segments spoken in isolation or in such short bits of speech as CVC monosyllables, etc. Conversational speech offers vast numbers of differing contexts and speaking conditions, which in turn causes the acoustic properties of the stream of speech sounds to vary, sometimes to an extreme degree. If an investigator wishes to synthesize a long utterance, it is most convenient to have a real speaker produce the speech, to analyze that sample for the appropriate acoustic parameters, and to use the measured parameters to produce the speech sample; in the method proposed here these would be the pronunciations of the same sentence by native and the foreigner.

Much more difficult, but theoretically more satisfying, is the technique of synthesis by rule (Holmes, Mattingly, and Shearme 1964; Rabiner 1969; Coker 1967). In this technique, the investigator seeks to establish the rules which underlie the way in which speech segments/syllables are conjoined in long strings. Such rules would predict the f_0, I_0 contours, the coarticulations, the relative durations of consonants, vowels, clusters, syllables, etc., and a host of other changes in segmental speech patterns characteristic of flowing speech and may be best realized not by using an acoustic-parameter controlled synthesizer, but by using a synthesizer which accepts articulatory commands, and in turn translates the commands to vocal tract movements/shapes, and thence to an acoustic signal (Coker 1967).

Natural Pronunciation of Speech

Let us suppose that naturally spoken speech produced by a proficient speaker of his native tongue is characterized by a vide variety of rhythmic, intonational, stress, spectral-segmental, and temporal parameters, etc. Within limits, these parameters are in large part dictated by the phonology and syntax of the speaker's native language, and in part by the speaker's personal characteristics; in other words (Ladefoged and Broadbent 1957), linguistic and paralinguistic factors influence the parameters of speech.

4. Testing the "Naturalness" of Pronunciation

In this section of the paper, we wish to present the thesis that speech synthesis offers an elegant and efficient way to determine what factors determine the naturalness of pronunciation of a second language by a second language learner. Let us argue that the naturally spoken, high-quality speech of a native speaker of Language

X, L_x, is characterized by a set of segmental Sg_x, coarticulatory C_x, syllabic Sy_x, and suprasegmental Su_x parameters such that

$$L_x \quad E \quad \left\{ \begin{array}{ll} Sg_x & x: 1, 2 \ldots n \\ C_x & x: 1, 2 \ldots m \\ Sy_x & x: 1, 2 \ldots l \\ Su_x & x: 1, 2 \ldots k \end{array} \right\}$$

We have no a priori way of knowing, for a given language, which of the four classes of acoustic cues or which combination of these classes is most important to the perceived "naturalness" of speech in that language. Nor, given one of the classes of cues, can we give any relative ranking of the primacy of the various acoustic cues within such a class. For that matter, there may be more or fewer than 4 major classes of acoustic cues.

The experimental paradigm for testing naturalness of speech production consists of the following steps. As step one, native speakers S_{nx} of Language L_x are recruited, and tape recordings are made of their productions of a wide variety of sentential material. It is our hypothesis that the average speech utterance is of clause, phrase, or sentence length; hence speech samples of that size should comprise the corpus of speech stimuli to be used. Following that, a set of speakers who are learning language X, L_x, as a second language, S_{2x}, will be recruited and asked to speak the same set of utterances spoken by the native speakers of L_x. This results in two corpora of utterance US_{nx}, US_{2x}, one by native speakers US_{nx} and one by second language speakers of L_x, US_{2x}.

Both sets of utterances, US_{nx} and US_{2x}, are then subjected to a thoroughly exhaustive acoustic analysis which results in an inventory of segmental, coarticulatory, syllabic and suprasegmental cues for each utterance in the corpus. Thus

$$US_{nx} \in \left\{ \begin{array}{ll} Utt_1 & \left\{ \begin{array}{l} Sg_{x1} \\ Sy_{x1} \\ C_{x1} \\ Su_{x1} \end{array} \right\} \\ Utt_2 & \\ \vdots & \\ Utt_n & \left\{ \begin{array}{l} Sg_{xn} \\ Sy_{xn} \\ C_{xn} \\ Su_{xn} \end{array} \right\} \end{array} \right\}$$

where each class of cues for an utterance is really a set of time-varying cue-values across the length of the utterance

In addition to the set of acoustic cues for each utterance in the corpus, for each speaker, it is to be hoped that by patient work and inspection, the entire set of acoustic cues of all types can be collapsed across utterances, U, and speakers, S_x, so that a final subset of acoustic parameters characteristic of L_x can be arrived at which can be used for control of a speech synthesizer. With this data in hand, the experimenters can then use US_{nx}s to synthesize a set of synthetic utterances which, it is hoped, will be a near match in quality, to the naturally spoken set of utterances US_{nx}. That is, by using L_x, we would hope to produce an artificial, synthesized set of utterances, $US_{nx_{synthetic}}$ whose quality is a good approximation to the quality of the naturally spoken set of utterances, $US_{nx_{natural}}$. Obviously, many attempts will have to be made at synthesizing utterances before a match of sufficiently good "perceptual quality" is made between the natural $US_{nx_{natural}}$ and synthetic $US_{nx_{synthetic}}$ utterances. This process will consist of using the set acoustic parameters measured for each utterance, $US_{nx}\left\{\begin{matrix} Sq_{xn} \\ Sy_{xn} \\ C_{xn} \\ Su_{xn} \end{matrix}\right\}$ as well as the condensed set of general acoustic parameters characteristic of L_x which have been distilled from close examination of each utterance of each native speaker. Recursive perturbation of the control parameters of such a synthesizer as the Klatt (Ibid.) device using the above parameters will then result in a high quality set of L_x utterances, $US_{nx_{synthetic}}$, which will be the stimuli for the perceptual testing.

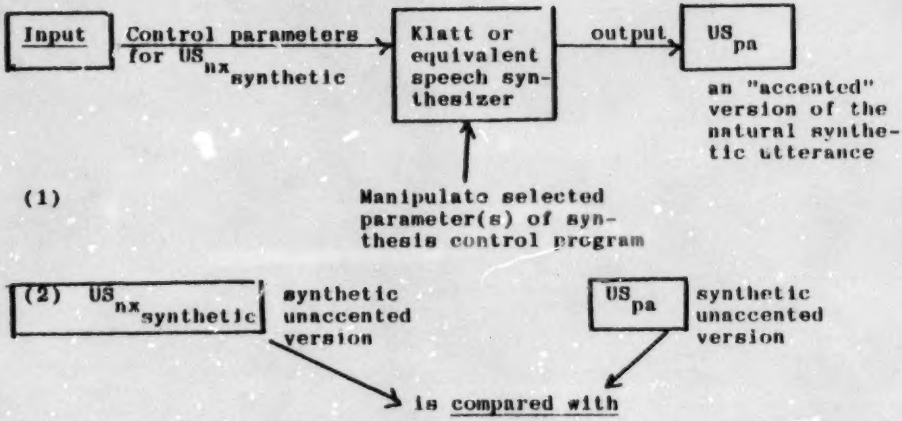
Perceptual Testing

The $US_{2x_{natural}}$ set of utterances of the second language learners will then be scrutinized acoustically and a set of acoustic parameters for each of their utterances will be extracted.

$$US_{2x} = \left\{ \begin{matrix} U_1 & \left\{ \begin{matrix} Sg_{2x_1} \\ Sy_{2x_1} \\ Su_{2x_1} \\ C_{2x_1} \end{matrix} \right\} \\ U_2 & \\ \vdots & \\ U_n & \left\{ \begin{matrix} Sg_{2x_n} \\ Sy_{2x_n} \\ Su_{2x_n} \\ C_{2x_n} \end{matrix} \right\} \end{matrix} \right\}$$

At this point, a close comparison can be made between the parameters of $US_{nx_{natural}}$ and $US_{2x_{natural}}$ to determine along what parameters of speech the second language learners differ most markedly from native speakers, S_x of L_x. This process will undoubtedly be most tedious, but it can be facilitated by use of published reports concerning the segmental, syllabic, coarticulatory, and suprasegmental characteristics of L_x, and of the native language of the S_{2x} speakers. It would be expected that the phonological competence of S_{2x} speakers in their native language, which we will call L_y, will interfere with their second language speech productions in ways predicted by t'Hart (1971).

Based on a comparison of acoustic cues in US_{nx} and US_{2x} speech productions in L_x, the experimenter could assemble a set of working hypotheses concerning which parameters of speech contributed most to the "accent" of the S_{2x} speaker as he attempts to speak in L_x. At this point the experimenter would first load the set of articulatory parameters into the Klatt synthesizer which produced a "natural" sounding synthetic utterance in L_x. Then the experimenter would systematically perturb the articulatory synthesis control program for that utterance and produce a "prototype-accented" U_{pe} version of that utterance. The experimental sequence would consist, in theory, of two steps, (1) synthesis and (2) perceptual testing, as shown in outline below



via any of a number of rating techniques in standard use. For example, any two utterances can be compared with each other in terms of their naturalness by

(1) pair comparison

(2) magnitude estimation, and

(3) non-parametric, multidimensional scaling

In essence, the synthetic accented utterance and the synthetic unaccented utterance are presented in pairs to native speakers of L_x. The native listener is then asked to judge the degree of difference in "quality" between the accented and unaccented synthetic version of the utterance. The process can be best illustrated by presenting a hypothetical example. Let us suppose that we are investigating the degree to which native speakers of English have mastered Spanish pronunciation. After a native speaker of Spanish has spoken a set of Spanish utterances, they are analyzed acoustically and after trial/error efforts, the derived set of acoustic parameters can be used to create synthetic versions of the native Spanish speakers' utterances which have a perceptual quality of production which matches the real utterances quite closely. The same set of utterances are spoken by native speakers of English learning to pronounce Spanish, and then analyzed acoustically. Scrutiny of the results as is described by t'Hart (1971) suggests to the experimenter that non-native pronunciation differs, let us say, most markedly from native production in that non-natives tend to show differential vowel

lengthening and diphthongization of vowels not shown by the native speaker. The experimenter then, using the control parameters for a high "quality" synthetic Spanish utterance deliberately introduces in a precisely known and controlled fashion, varying amounts of diphthongization and vowel lengthening both separately and in combination. Thus, an entire ensemble of synthesized utterances with potential "accent" are created. These stimuli are then randomized and paired with the synthesized, unaccented versions of that utterance derived from the speech of the native speakers of Spanish. These are then played to a crew of listeners, all native speakers of Spanish, who are asked to rate each pair in terms of how much more natural or unnatural the synthetic accented utterance is than the synthetic unaccented utterance. Using multidimensional scaling procedures, the experimenter can then determine the degree to which various degrees of diphthongisation alone, vowel lengthening alone, or various combinations of both articulatory parameters cause a synthetic utterance to sound as if it had a non-native "accent". The virtue of this method is that it is quantifiable and controllable. Any single articulatory parameter, or any combination of parameters can be varied in a step-wise fashion to determine how they singly or conjointly contribute to the perception of "accent" in the speech of second language learners. By use of regression analysis and multidimensional scaling procedures, a large ensemble of acoustic parameters can be tested for their solitary and joint contribution to perceived "accent", and in fact, a "weighting" factor can be assigned to each parameter in terms of its relative contribution to the perception of second-language "accent". If such a program of systematic research were introduced, within the decade investigators could provide teachers of foreign language with a list of articulatory-perceptual parameters crucial to "natural" accent. The teachers could then implement these data by placing special emphasis on the crucial parameters during training.

Drawbacks of the method

There are a number of technical and methodological drawbacks to the use of A x S techniques to determine what factors are pre-potent in causing a second language learner's speech to sound "non-native" in quality. First, a fully equipped digital-speech synthesizer demands great skill in computer programming, and the use of a modern mini-computer fully equipped with sufficient memory, mass storage, I/O devices, and soft-ware to make high quality synthesis feasible. These items of hard and software are not only expensive, but demand the time and effort of skilled and dedicated programmers. Getting such a system up and functioning is a tedious task. Second, as good as digital synthesizers such as the Klatt synthesizers are, they are insufficiently flexible and elegant to produce synthetic speech whose "naturalness" is undistinguishable from that of "real" speech. Perhaps the single greatest difficulty is the problem of trying to synthesize "long", multi-syllabic utterances which sound sufficiently natural to serve as stimuli for such a study. Mattingly (1968) cites the difficulties inherent in trying to synthesize long utterances

by use of General Rules of Synthesis. In practice, given an utterance, one would like to be able to synthesize the utterance by knowing only the phonemic-segment string used in the utterance. The experimenter, in theory, would have available a dictionary of acoustic parameters for the "ideal" target production of each phoneme-segment, or extrinsic allophone, in the utterance. A series of general-conjoining rules would tell him, automatically, how to vary the segment parameters to create coarticulated syllables, how to vary stress, how to create an intonation contour, how to insert junctures, and how to vary segmental, syllabic, and phrasal durations to get an appropriate rhythm and rate to the speech. We are really very, very far from having such an operational program. Rather than using a set of elegant, hierarchical rules to generate natural utterances of realistic length, we are generally reduced to having a speaker record the utterance, which is then analyzed spectrographically, and a set of time-varying acoustic control parameters are extracted for the whole utterance, and inserted in the synthesizer, which then synthesizes the utterance in a stepwise fashion whose smoothness depends upon how swiftly control parameters can be changed across time. Such an utterance inescapably, even if synthesized, bears the strong idiolectal imprint of the speaker who produced the template utterance from which the control parameters were extracted. Nor, in such a case, can the investigator easily decide on how to "naturally" shift the tempo, rhythm, stress, emphasis, juncture, etc. of such a synthesized utterance since no generalized rules of synthesis were used to synthesize it in the first place. Equally potentially frustrating is that two different native speakers of a language may use somewhat differing and idiosyncratic rules of articulatory motor control to produce utterances which sound equally "natural" for differing reasons. Using our experimental paradigm, we would arrive, undoubtedly, at a somewhat different weighting of speech parameters which contribute to "naturalness" of speech depending on which native speaker was used for initial extraction of control parameters to drive the speech synthesizer. Hopefully, most of the parameters of speech which make it sound "natural" in a given language are part of the linguistic competence of all native speakers, and would vary only slightly because of idiosyncracies and performance constraints between speakers.

These caveats aside, it is our opinion that the Analysis x Synthesis technique is a most promising means of investigating the problem of how to determine what causes speech to sound more or less "natural" when spoken by second language learners.

REFERENCES

Celani, Maria Antonieta. Phonology at the teacher-training level. English Teaching: a regular bulletin for teachers of English in Brazil (Rio de Janeiro), 2, 10-14, 1967.

Coker, C. Synthesis by rule from articulatory parameters. Proceedings of the 1967 Conference of Speech Communications. Process. Paper A9, 52-53, 1967.

DeGrève, Marcel. New methods of teaching foreign languages. Revue des Langues Vivantes (Brussels), 34, 79-93, 1968.

Dimitrijević, Naum R. and Djordjević, Dušan. The reliability of the subjective assessment of the pupils' pronunciation of English as a foreign language. IRAL, 9, 3, 245-65, 1971.

Fant, G. (ed.) Proc. Speech Comm. Seminar: Stockholm. April 3, 1974. Vol.2 "Speech Production and Synthesis by Rules." Stockholm: Almqvist & Wiksell. 1974.

Fant, G. Acoustic Theory of Speech Production. Mouton: The Hague. 1960.

Flanagan, J.L. Note on the design of "terminal analog" speech synthesizers. J. Acoust. Soc. Amer., 29, 306-310, 1957.

Flanagan, J.L. and Rabiner, L.R. Speech Synthesis. Bench-mark Press (Wiley). 1973.

Fries, Charles C. Teaching and Learning English as a Foreign Language. Ann Arbor: Univ. Michigan Press. 1945.

Holmes, J.N., Mattingly, I. and Shearme, J.N. Speech synthesis by rule. Lang. & Speech, 7, 127-143, 1964.

Klatt, D. A cascade/parallel terminal analog speech synthesizer and a strategy for consonant-vowel synthesis. Paper presented to the 92nd meeting, Acoustical Soc. America. State University, Pennsylvania. Spring 1977.

Ladefoged, P. and Broadbent, D.E. Information conveyed by vowels. J. Acoust. Soc. Amer. 29, 98-104, 1957.

Lado, Robert. Linguistics Across Cultures: Applied Linguistics for Language Teachers. Ann Arbor: Univ. Michigan Press, 1957.

Lebel, Jean Guy. The contribution of systematic exercises in auditory discrimination for phonetic correction. Revue de phonétique appliquée, 10, 3-9, 1969.

Mattingly, I.G. Synthesis by rule of G.A. English. Status Report on Speech Res. Haskins Labs. 1968.

Rabiner, L.R. 1969. A model for synthesizing speech by rule. IEEE Trans. Audio Electroacoust. AU-17, 7-13.

Reichmann, Eberhard. Some visual aspects of pronunciation. German Quarterly, 40, 3, 398-403, 1967.

Renard, Raymond. On the tension factor in phonetic correction. Revue de phonétique appliquée, 17, 71-86, 1971.

Scholes, R.J. The categorization of synthetic speech sounds as a predictive device in language teaching. Journal of English as a Second Language, 2, 2, 1-16, 1967.

Scholes, R.J. Phonemic interference as a perceptual phenomenon. Lang. & Speech, 11, 86-103, 1968

Studdert-Kennedy, M. Speech production. Chapter in N. Lass, ed. Contemporary Issues in Experimental Phonetics. Academic Press: New York. 1976.

t'Hart, J. A note on linguistic implications of second language phonetics. In L.L. Hammerich, R. Jakobson, E. Zwirner, eds. Form and Substance: Phonetic and Linguistic papers to Eli Fischer Jørgensen. Akademisk Forlag: Copenhagen. 141-148, 1971

Trocme, Helene. Difficulties of pronunciation experienced by French learners of English. IUT Bulletin Pedagogique, Paris,30, Langues Vivant 19, 28-35.